

score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

Declaración

AY593927 Se

АВІАЦІА

AC073054 Use

ADONIS 0013790X9200045A

ADDITIONAL INFORMATION

RECORDED
INDEXED
FBI

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED

AL010030 NC

AL06820 / MC

AP006305 HC

AL591164 HC

AR164611 HC

AC024626 HC

AC006035 HC

AC009427 HC

78007227 Hc

AC074117 Use

●

000000000000

AL031348 I

AL596188 NC

AC124944 HC

Drafting

10/11/2011

10,000.

FEATURES	CHIRON CORPORATION (US)
SOURCE	location/Qualifiers 1..879 /organization="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"
ORIGIN	

Query Match	100.0%	Score	879	DB 6	length	879	
Best Local Similarity	100.0%	Pred. No.	8	8e-263			
Matches	879	Conservative	0	Mismatches	0	Indels	0
						Gaps	0

QY	1	TGTGGGGAAAAGCAGAAGATCAGATTGTATCTGTCTGTGTGAAGAAGATAGACAT	60
Db	1	TGTGGGGAAAAGCAGAAGATCAGATTGTATCTGTCTGTGTGAAGAAGATAGACAT	60
QY	61	AGGAGACTCCATTTTGTATGTACTAAGAAAAATCTTCTGCTTGAAGATTCGTTTAATC	120
Db	61	AGGAGACTCCATTTTGTATGTACTAAGAAAAATCTTCTGCTTGAAGATTCGTTTAATC	120
QY	121	TATGACCTTACCCCAACCCCGTGCTCTGAAACAATGCGCTGTGTCCACTCAGGGTTAA	180
Db	121	TATGACCTTACCCCAACCCCGTGCTCTCTGAAACAATGCTGTGTGTCTCACTCAGGGTTAA	180
QY	181	ATGATTTAAGGGCGGTGACAGATGTGCTTTTGTAAACAGATGCTTGAAGCAGATGCTC	240
Db	181	ATGATTTAAGGGCGGTGACAGATGTGCTTTTGTAAACAGATGCTTGAAGCAGATGCTC	240
QY	241	CTTAAAGATCATCACCCTCTTAAATCTTCACAGTACCACGAGACACAAAACTGCGGAAAG	300
Db	241	CTTAAAGATCATCACCCTCTTAAATCTTCACAGTACCACGAGACACAAAACTGCGGAAAG	300
QY	301	CCGCAGGACCTCTGCTTAGGAAACCCAGATATTGTCCACGTTTCTCCCATGTGATAG	360
Db	301	CCGCAGGACCTCTGCTTAGGAAACCCAGATATTGTCCACGTTTCTCCCATGTGATAG	360
QY	361	CCTGAATATAGGCTCTGTGGGAAAGGAAAGACCTGACCGTCCCCACGCCGACACCCGTA	420
Db	361	CCTGAATATAGGCTCTGTGGGAAAGGAAAGACCTGACCGTCCCCACGCCGACACCCGTA	420
QY	421	AAGGATCTGTGCTGTAGAGAGATTAGTTAAAGAGAAAGAAATGCTCTTGACAGTTAGACA	480
Db	421	AAGGATCTGTGCTGTAGAGAGATTAGTTAAAGAGAAAGAAATGCTCTTGACAGTTAGACA	480
QY	481	AGAGAAAGCAATCTGTCTCTGCTGCTGCTGCGGCAATGAAATGTCTGTGTAAACCC	540
Db	481	AGAGAAAGCAATCTGTCTCTGCTGCTGCTGCGGCAATGAAATGTCTGTGTAAACCC	540
QY	541	GATTGTATGCTCCATCTACTGAGATAGGGAAGAAACCGGCTTAAGGGCTGAGAGGTGGACCT	600
Db	541	GATTGTATGCTCCATCTACTGAGATAGGGAAGAAACCGGCTTAAGGGCTGAGAGGTGGACCT	600
QY	601	GCGGCGACCAATACTGCTTGTGTAAAGCACTGAGATGTTATGTATGATCATATCTTAAAG	660
Db	601	GCGGCGACCAATACTGCTTGTGTAAAGCACTGAGATGTTATGTATGATCATATCTTAAAG	660
QY	661	CACAGCACTTAAATCCTTTAACAATGTGTATGATGCAAGAACCTTGTGTCAATGTTGTCT	720
Db	661	CACAGCACTTAAATCCTTTAACAATGTGTATGATGCAAGAACCTTGTGTCAATGTTGTCT	720
QY	721	GCTGACCTCTCCACAAATTTGTCTTGTGACCTTGACACATCCCTCTTGTGAGAAACAC	780
Db	721	GCTGACCTCTCTCCACAAATTTGTCTTGTGACCTTGACACATCCCTCTTGTGAGAAACAC	780
QY	781	CCACAGATGATCAGTAAATACTTAAGGGAACCTCAGAGGCTGAGCGGGAATCTCATATGCTG	840
Db	781	CCACAGATGATCAGTAAATACTTAAGGGAACCTCAGAGGCTGAGCGGGAATCTCATATGCTG	840
QY	841	AAAGCTGAGTCCCGGGGTCCCTTCTTCTTCTTCTCATAT 879	
Db	841	AAAGCTGAGTCCCGGGGTCCCTTCTTCTTCTTCTCATAT 879	

RESULT 2				
AX594072				
LOCUS	AX594072	968 bp	DNA	linear
DEFINITION	Sequence 150 from Patent WO0246477.			PAT 13-FEB-2003
ACCESSION	AX594072			
VERSION	AX594072.1	GI:28375305		
KEYWORDS				
SOURCE				
ORGANISM	Homo sapiens (human)			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1			
AUTHORS	Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.			
TITLE	Endogenous retroviruses up-regulated in prostate cancer			
JOURNAL	Patent: WO 0246477-A 150 13-JUN-2002;			
	CHIRON CORPORATION (US)			
FEATURES				
source	location/Qualifiers			
	1..968			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:9606"			
ORIGIN				

Query Match	100.0%	Score 879;	DB 6;	Length 968;
Best Local Similarity	100.0%	Pred. NO.8.9e-263;		
Matches 879;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	1	TGTGGGGAAAAAGCAGAGAGATCAGATTGTATCTGTCTGTGTGAGAAAGAAAGATGACAT	60
Db	1	TGTGGGGAAAAAGCAGAGAGATCAGATTGTATCTGTCTGTGTGAGAAAGATGACAT	60
OY	61	AGGAGACTCCATTTTGTATGTACTAGAAAAAATCTCTTGCTTGAGATTCTGTATATC	120
Db	61	AGGAGACTCCATTTTGTATGTACTAGAAAAAATCTCTTGCTTGAGATTCTGTATATC	120
OY	121	TATATACCTTACCCCCCAACCCCGTCTCTCTGAAAACATGTGTCTGTGTACATCTCAGGGTTAA	180
Db	121	TATATACCTTACCCCCCAACCCCGTCTCTCTGAAAACATGTGTCTGTGTACATCTCAGGGTTAA	180
OY	181	ATGTGATTAAAGGGCGGTGACAGATGTGCTTTGTTTAAACAGATGCTTTGAAAGCAGACATCTC	240
Db	181	ATGTGATTAAAGGGCGGTGACAGATGTGCTTTGTTTAAACAGATGCTTTGAAAGCAGATCTC	240
OY	241	CTTTAAGATCATACACACTCCCTTATCTCAAGTACCAGGAGACAAAATCTGCGAAGG	300
Db	241	CTTTAAGATCATACACACTCCCTTATCTCAAGTACCAGGAGACAAAATCTGCGAAGG	300
OY	301	CCGAGAGGACCTCTGCTTAGAAAAAGCCAGTATTTGTCCACGTTTCTCCCATGTGATAG	360
Db	301	CCGAGAGGACCTCTGCTTAGAAAAAGCCAGTATTTGTCCACGTTTCTCCCATGTGATAG	360
OY	361	CCTGAATATATGAGCTCGTGGGAAAGGAAAGACCTGACCGTCCCCAGCCGACACCCGTA	420
Db	361	CCTGAATATATGAGCTCGTGGGAAAGGAAAGACCTGACCGTCCCCAGCCGACACCCGTA	420
OY	421	AAGGTCCTGTCTGAGAGAGATTAGTAAAAAGAGAAAGAAATGCTCTTGACATTGAGACA	480
Db	421	AAGGTCCTGTCTGAGAGAGATTAGTAAAAAGAGAAAGAAATGCTCTTGACATTGAGACA	480
OY	481	AGAGAAAGGACATCTGTCTCTGCTGTCTCCTTGAGGCAATGGAATGTCTTCGGTATPAAACC	540
Db	481	AGAGAAAGGACATCTGTCTCTGCTGTCTCCTTGAGGCAATGGAATGTCTTCGGTATPAAACC	540
OY	541	GATTGTATGCTCCACTACTAGAGATAGAGAAAAACGCTTAGGGCTGAGAGTGGAGCT	600
Db	541	GATTGTATGCTCCACTACTAGAGATAGAGAAAAACGCTTAGGGCTGAGAGTGGAGCT	600
OY	601	GCGGGACAGCAATACGTCTTTGTAAAGCACTGAGATGTTTATGTGTATGCAATATCTAAAG	660
Db	601	GCGGGACAGCAATACGTCTTTGTAAAGCACTGAGATGTTTATGTGTATGCAATATCTAAAG	660
OY	661	CACAGCACTTATCTTTACATTGTCTATGTATGCAAGACCTTTGTTCACATGTTTGTCT	720
Db	661	CACAGCACTTATCTTTACATTGTCTATGTATGCAAGACCTTTGTTCACATGTTTGTCT	720

```

Db      661 CACAGACCTTAATCTTTAATGCTTAATGCAAGACCTTTGTTCAATGTTTCT 720
Qy      721 GGTGACCTCTCCACACAAATGCTTGAACCTGACACATCCCTCTTGAGAAAC 780
Db      721 GGTGACCTCTCCACACAAATGCTTGAACCTGACACATCCCTCTTGAGAAAC 780
Qy      781 CCAAGATGATCAATAATACTAAGGAACCTGAGGCGGAGATCTTCATATCTG 840
Db      781 CCAAGATGATCAATAATACTAAGGAACCTGAGGCGGAGATCTTCATATCTG 840
Qy      841 AAGCGTGTGTTCCCGGGTCCCTCTTTCTTTCTCTATA 879
Db      841 AAGCGTGTGTTCCCGGGTCCCTCTTTCTTTCTCTATA 879

RESULT 3
AF164614 9472 bp DNA linear PRI 03-SEP-1999
LOCUS      Homo sapiens endogenous retrovirus HERV-K108, complete sequence.
DEFINITION
ACCESSION AF164614
VERSION    AF164614.1 GI:5802820
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 9472)
            Barbuiescu,M., Turner,G., Seaman,M.I., Delnard,A.S., Kidd,K.K. and
            Lenz,J.
            Many human endogenous retrovirus K (HERV-K) proviruses are unique
            to humans
            Curt. Biol. 9 (16), 861-868 (1999)
TITLE      Journal
MEDLINE    99400989
PUBMED     10469592
REFERENCE  2 (bases 1 to 9472)
            Barbuiescu,M., Turner,G., Seaman,M.I., Delnard,A.S., Kidd,K.K. and
            Lenz,J.
            Direct Submission
            Submitted (02-JUL-1999) Molecular Genetics, Albert Einstein College
            of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
FEATURES
            source
            1..9472
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                1..9472
                    /rpt_family="human endogenous retrovirus HERV-K108"
                    1..968
                        1112..6749
                            /gene="gag-pro-pol"
                            /gene="gag-pro-pol"
                            /join(1112..3108,3108..3915,3915..6749)
                            /note="-1 frameshifts presumed to occur within the codons
                            for the last amino acids in the Gag and Pro open reading
                            frames"
                            /codon_start=1
                            /product="Gag-Pro-Pol protein"
                            /protein_id="AAD51797.1"
                            /db_xref="GI:5802821"
                            /translation="MGOTKSKIKSKVASYLSFIKILKRGGVKSTNLIKLFQIIBO
                            FCPMPFGTLDLKDMDRIKELKELKQAKKNITPLTVMWMAIKALFEPOMEDSV
                            SVSDARSCITDCNENTRKSKOKETBELHEVYAEFMASTONVDYNOUEVYPT
                            LKLEGPVELVGPESKRPGRAPVPOVPTLPOKQOVENKTPQVAYOYMPFAL
                            QYRPPESQYQPGMPAPQGRAPVPOVPTLPOKQOVENKTPQVAYOYMPFAL
                            TEAWOPVTELEMPGEGAGGEPVTEAYKFSIKLMDKMGVQYQPNPYMT
                            LBSIAHGRLIPYDMEILAKSSLSPOQFTWIDGVOEYVRRAANPYNIDA
                            DOLGIGONSTISQALMONEATEIOVRATCLAMEKIDOBSTCPSEFNTVROSKEP
                            YRPFARLOUVAOKSIADKARKIYIVLAEYANPCQSAIKYIKLVAGSVISE
                            YVACDGTIGAMKAMLAONITGVLAGOVRTGRKCTNCGOIGHLKACPVLNKN
                            ITTQATTGRBPDLPCRCKKGRMAQCSKSKEDKNGQPLSGNDGROQAPQDTGA
                            PIOPVPOGPOGQPLSOVFOGISOPLPNNCPPOAVOVDLCTIOVLSIGBP
                            POKPTGVGPPLKGTGLIGRSNLNKGQITSVSDYKGIOLVSSISIPMA
                            SPDRIAQLLLPYIKGNSEIKRIGIGSTDPTRKAAYASQVSENRPFYKAIIOGK

```

```

QFEGLVDTGADVSIILNMPKMKPKQKAVTGVIGTASBEVYOSTEILHCLBPDS
STVQPMITSIPLNLMGRIQLQWGAETIMPAISPSQKIMTKMGIPEKGIKND
GKIPEAKINOREGIDNCLGATAPPEKPIPLTKTEKPAVYVNNWNPYKQLEAL
HILANOLEKGIHTEPSPNNSPVFVIOKSGKRMITDIRAVAVIOPMPGPIB
SPAMIPKDMPLIITDKOCFTPIABODEKFAFTPIANNKPAIRFOWKUTPOG
LNSPTTCOTVVGALDQVREKFSDCYIHICIDILCAEFTDKLIDCTYLOAVNA
GLIASDKIOTSPPFHYLMQJENRKIKPKKPIKIDOTLTLNDFOKLADIMVIRP
LGIPTAMSNLFSILRGSDNSKRMTPATKEIKVTEKIQSAQINRIDPLAPLO
LIPATASPTGIIIONTDLYMSFTPLSTYKTFPLVYDQATLIGOTRLIKLGD
PDKIVPLTEQVROAFINSKAMKIGIANVGIIDNVPRTKIFQELKLTWILPKIT
RREPLENALTVFTDGSNGKAAATGPERKIKTPYQSAQAEVAVYVUODPOPI
IISDSANVOATADVETALIKYSMDQNLQFNLQOTVRKRNPFYTHIRAHINQ
GPLTKANEQADLVSSALIKAOEHLALTHVNAAGLNKRPVTKROAKDIDQINQ
LHLPOTBAGVNPGLCPNALMOMDVTHVPSPGRLSYVHVTVDYSHFIYATOCGSE
SHVKKHLSCFAYMGVPEKIKTDNGPBGCSKAFQFLSOKKISHTSIPIYNSQOAV
ERTNRITKPOLVQKQEGDSKECTTPMOANLALYTNPLNRYNNTTSABOHLCK
KNSPHBGLIWMKDNKKTVEIGVITVWGFQSGSGENQLPWVITPRLKFTNEI
RDAKSTSAETETSSSTVDSODEQNDVARTBEVALHOGRAANLIGTLEADAVSYK
ISREHKDTPREYAAACSLDDCINGKSPYACRSSCS"
6451..8550
/gene="env"
6451..8550
/codon_start=1
/product="envelope protein"
/protein_id="AAD51798.1"
/db_xref="GI:5802822"
/translation="MNPSEMQKAPRRRRHRRAPIPTHQNKVTSBEQMLPSTK
AEPVTAQLKLQLATKYLENKKVTPSPVMIADRCVAKSEEMNINSIGYH
YMAVPEPPIRAVATMDNTEVYVNDSPWPGIDIRCAKSEEMNINSIGYH
PPICGAPGCLMPAVQNMVLEVPTVSPICRFYMWGSMGLPRVYLDQFOSRSL
KPIPKGPKCKEIKPKSKNTEVLMVECSVANSAYILONNEFTGIIIDAPGQFYHNS
GOROSGCSAOVSPAVSDLTLESIDKHKHKLQSLSPYMEKESKISTEPRIYVSGP
EHPBLRITVASHIRIRKSNQTLERDRKPFITIDNSLITPLQSCVAPRYLVYG
NIVIKDSOTITTCENCLTICIDSTEFWQRIILVRRBEVMIPEVSDREKESPVH
ILTEVKGIVNRSKRFIFTLIAVIMGLIATVADAAVAGVLAHSVQSVNVNMQNS
TRLMSQSSIDOKLANQINDLROTIVMGRLNLSLEHFPOLQCDMNTSDCITPOLYN
ESEHNDMVRHLOGREDNLTDISKEDQIFESKHLNLVPTGETAIVAGVADLANL
NPYTWYTGSTTITINILIVCLFCLLLVCRCTQQLRRSDHREARMVMAVLSRK
GNVGVSKRQIYTVSY"
8504..9472
ORIGIN
LTR
Query Match      100.0%; Score 879; DB 9; Length 9472;
Beat Local Similarity 100.0%; Pred. No. 1,1e-262;
Matches 879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTAGAAAAGTAGACAT 60
Db      1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTAGAAAAGTAGACAT 60
Qy      61 AGGAGACTCATTTTGTATGACTTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
Db      61 AGGAGACTCATTTTGTATGACTTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
Qy      61 AGGAGACTCATTTTGTATGACTTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
Db      61 AGGAGACTCATTTTGTATGACTTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
Qy      121 TATGACCTTACCCCAACCCCGCTCTGAAAACAGTGTGTGCACATCAGGGTTAA 180
Db      121 TATGACCTTACCCCAACCCCGCTCTGAAAACAGTGTGTGCACATCAGGGTTAA 180
Qy      181 ATGATTAAGGGCGGTGACAGATGTGCTTTGTTAAAGATGCTTGAAGGACAGATCTC 240
Db      181 ATGATTAAGGGCGGTGACAGATGTGCTTTGTTAAAGATGCTTGAAGGACAGATCTC 240
Qy      241 CTTAAGAGTATCAACAATCTTAATCTCAAGTACCAAGGACCAAAAACTGCGAAG 300
Db      241 CTTAAGAGTATCAACAATCTTAATCTCAAGTACCAAGGACCAAAAACTGCGAAG 300
Qy      241 CTTAAGAGTATCAACAATCTTAATCTCAAGTACCAAGGACCAAAAACTGCGAAG 300
Db      241 CTTAAGAGTATCAACAATCTTAATCTCAAGTACCAAGGACCAAAAACTGCGAAG 300
Qy      301 CCGAGAGGACCTGTGCTTAGAAAACCAAGTATGTCACGTTTCCCCATGATAG 360
Db      301 CCGAGAGGACCTGTGCTTAGAAAACCAAGTATGTCACGTTTCCCCATGATAG 360
Qy      361 CTTGAAATATAGCTCTGTGGGAAAGGAAACCTGACCGTCCCGACCGGACACCGGTA 420
Db      361 CTTGAAATATAGCTCTGTGGGAAAGGAAACCTGACCGTCCCGACCGGACACCGGTA 420

```

Db		361	CCTGAAATATGGCCTCGTGGGAAGGAAAAGACCTGACCCTCCCAAGCCGACAACCCGTA	420
Oy		421	AAGGTCCTGTCTCTAGAGAGATTAGTAAAAAGAGAAGATGCTTTTGACGTTGAGACA	480
Db		421	AAGGTCCTGTCTGAGAGATTAGTAAAAAGAGAAGAAATGCCCTTTCAGTTGAGACA	480
Oy		481	AGAGAAAGGCATCTGTCTCCCTGCTCCCGGCAATGAAATGTCCTCGATATAAACCC	540
Db		481	AGAGAAAGGCATCTGTCTCCCTGCTCCCGGCAATGAAATGTCCTCGATATAAACCC	540
Oy		541	GATTGTATGCTCATCTACTAGATAGGGA AAAA CCGCTTAGGGCTGGAAGTGGAGCT	600
Db		541	GATTGTATGCTCATCTACTAGATAGGGA AAAA CCGCTTAGGGCTGGAAGTGGAGCT	600
Oy		601	GCGGCGAGCAAATACTGCTTTGTAAAGCACTAGATGTTATGTATGATATATTAAG	660
Db		601	GCGGCGAGCAAATACTGCTTTGTAAAGCACTAGATGTTATGTATGATATATTAAG	660
Oy		661	CACAGCATTTAATCCTTTACATTTGTCTATGTATGAAAGACCTTTGTTCACATGTTTGTCT	720
Db		661	CACAGCATTTAATCCTTTACATTTGTCTATGTATGAAAGACCTTTGTTCACATGTTTGTCT	720
Oy		721	GCTGACCTCTCCCCACAAATTTGCTCTTGAGACCTTGACATCCCTCTTTGAGAAAAC	780
Db		721	GCTGACCTCTCCCCACAAATTTGCTCTTGAGACCTTGACATCCCTCTTTGAGAAAAC	780
Oy		781	CCACAGATGATCAGTAAATACTAAAGGAACCTCAGAGCTGGCGGAGTCTTCATATGCTG	840
Db		781	CCACAGATGATCAGTAAATACTAAAGGAACCTCAGAGCTGGCGGAGTCTTCATATGCTG	840
Oy		841	AACGCTGGTCCC CGGGCTCCCTCTTCTTTCTCTATA	879
Db		841	AACGCTGGTCCC CGGGCTCCCTCTTCTTCTCTATA	879
RESULT 4				
AF074086		19221 bp	DNA linear	PRI 11-MAR-2002
LOCUS				
DEFINITION	Homo sapiens tandemly repeated human endogenous retrovirus HERV-K (HML-2.HOM), complete sequence.			
ACCESSION	AF074086			
VERSION	AF074086.2	GI:9558700		
KEYWORDS				
SOURCE				
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE AUTHORS	Mayer,J., Sauter,M., Raczy,A., Scherer,D., Mueller-Lantzsch,N. and Meese,E.			
TITLE	An almost-intact human endogenous retrovirus K on human chromosome 7			
JOURNAL MEDLINE	Nat. Genet. 21 (3), 257-258 (1999)			
PUBMED	99178257			
REFERENCE	10080172			
REFERENCE AUTHORS	2 (bases 1 to 19221) Reus,K., Mayer,J., Sauter,M., Scherer,D., Muller-Lantzsch,N. and Meese,E.			
TITLE	Genomic organization of the human endogenous retrovirus HERV-K(HML-2.HOM) (BRVK6) on chromosome 7			
JOURNAL MEDLINE	Genomics 72 (3), 314-320 (2001)			
PUBMED	21295053			
REFERENCE	11401447			
REFERENCE AUTHORS	3 (bases 1 to 19221) Mayer,J., Mueller-Lantzsch,N. and Meese,E.U.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-JUN-1998) Institut fuer Humangenetik, Universitaet des Saarlandes, Homburg 66421, Germany			
REFERENCE	4 (bases 1 to 19221) Reus,K., Mayer,J. and Meese,E.U.			
REFERENCE AUTHORS	Direct Submission			
TITLE	Submitted (25-JUL-2000) Institut fuer Humangenetik, Universitaet des Saarlandes, Homburg 66421, Germany			

REMARK	Sequence update by submitter
COMMENT	On Jul 28, 2000 this sequence version replaced gi:4456987.
FEATURES	Location/Qualifiers
source	1. .19211 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="7" /clone="Y20-B19" /clone_1fb="Lawrence Livermore National Laboratory Single Chromosome Library: IL07N0C1" 1. .11976 /note="HML-2, H0W, tandemly repeated" /rpt_family="Human endogenous retrovirus HERV-K" /rpt_type=dispersed 1. .968 971. .988 /note="lYaine primer binding site" 1112. .3112 /gene="gag" 1112. .3112 /gene="gag" /product="gag protein" /protein_id="AA88165.1" /db_xref="GI:9558702" /translation="MQGTSKIRKSYASVLSFKILLLKRGVRSVTKNLIKLFQIIQ FCWFEPOQTLDLDMKRIIGKELKQGRKNNIPLTVMDMAIILKALBEPQEBDSV SVSDAPGSCIIIDCENNRKSKQETEGHLEVEAPVMAQSTONVDNLOEVIYET LKKEGKPELVGSESKPRTGSLPAGVPTVLTQPOKQVKNTEPTQVWPPMEL QYRPPESQGYGPGMPAPGGRAPYPQPTRRRLNPTAPRQSGKTHETIDKRSKD TEMQPPTVLEPMPGSGAEGEPVTEAKTSFSIKGLDMKEGKQVQIPNSPYHRT LDSILAHGRLLPYDWEILAKSSLSPQQLFTWIMDVGVEQVRRRRANRPVNIDA DOLGIGQNNSTISQALMONEAIEQVRAICLAMEKIDPGSTCSFNTVRQSGKEP YVPEVARLDQVAOKSIADKARIVLEIAYENANPECCQAIKPLKGVAGSDVSE YVACRIGGAMKHAMMAQITGVVIGGVGFRRGRCAVCGOIGHLKNKCPVLNKN ITITQATTGREGPDLCPRCCKGKGMAQCSKRDKNQPLSGNBQGRQAPQDTAF PIQPFVQGFQGGQPLSVQFGGISQLPQNNCCPPQAAVQO" 42915. .3918 /note="translated by frameahift" /codon_start=3 /product="protease" /protein_id="AA88166.1" /db_xref="GI:9558703" /translation="WATIVYKRAKGPASGPTTNMGIPNSAICSSGSGTTPVPSVS GNKPVTTIQQLSPATSGSAVVDCTIOAVSLBGEPOKPTVYGLPKGTGLIG RSSILNKGVQIHSTVSDSDYKGEIOLVSSIPMSAPRRIQLLPLPIKGNSEI IKRIGSGPTPKGAAYMAVSQENRPNVCAITIGQFQEGVLTGADVSIILANQPK NMPEKQAVTGLVIGTRASEVYQSTETILHCLGPNQSTVQPMHTSIPLANMGDILLQO WGAHITMPAPSPPTSQIKMTKMGYIPGKGLGNEDGIKIPBAKINQEREGLNFC" 43878. .6749 /gene="pol1" 3878. .6749 /gene="pol1" /note="translated by frameahift" /codon_start=2 /product="polymerase" /protein_id="AA88167.1" /db_xref="GI:9558704" /translation="NNSKRKRNRBSLGAATVPPKPIPLTWKTEKRVWNNQPLPKQ KLEHLANLQLEKGIHPSFSPPANSPVVITQKSKMMLTDLRVANVITQMGEL QPGIPSPAMI PKOMPLIIDLKDCPFTPLABEDCECFATPTIPIANKKEAPFOMVY LPOGLNPSPTICQTFVBARLOPAREKFSDCYIHCIDIDILCAETKXKLDICVTFLOA EVANAGLAISDKIQTSTPHYIGMOENKRIKEIKRQITKLTNDQKGLGDIN WISPTGLPYTAMSNFSLTRGDSNDNSKMLPEAKKEIKLYEKKISQOINRIDL APLDLIFATASHPTGCIITNTDLVRSFPLPHSTVKTFITLYDQIATLTCQYLRIRIK LQGDQPKIVPLTKEGRQALNSGAMKGLANPVGIIDNHPKTKI FOPLKLTWII LPIKTRPELENALTFTDGSNGKAYTGPKEKRVIKTPYQSAQRAELVAVITVLODF DOPNIIISDAVYAOATRVETALIKYSMDQDLNOLFNLQCTVRKKNPFYITHIAPA HTNLPGLTANQADOLVSSALIKQELHATHVNAAGKKNEDVTWKOAKDIVQHC TQCOVAILPQEGANPRGLCPNALMDQVTVKPSGRSLSVYATVDTSHGFIATQO TGSTSTVNTKHLISCRAMGVPEKIRITNDNGPVCYCSAFOFSLQWKISHTTGYPSNO GQALIVETNNTLTQLVKQEGSDSEKCTPQOMLNALATTLNPLNRYNQTTTSAQO


```

Db      721 GGTGACCCCTCCCAACATTGCTTGTGACCTGACATCCCTCTTCGAGAAACAC 780
Qy      781 CCACAGATGATCAGTAATAATCTAAGGAACTCAGAGGCTGCGGATCTCCATATGCTG 840
        |||
Db      781 CCACAGATGATCAGTAATAATCTAAGGAACTCAGAGGCTGCGGATCTCCATATGCTG 840
Qy      841 AACGCTGTTCCCGGCTCCCTCTTCTTCTCTCTATA 879
        |||
Db      841 AACGCTGTTCCCGGCTCCCTCTTCTTCTCTATA 879
        |||

RESULT 5
AC072054/c      154037 bp      DNA      linear      PRI 08-OCT-2003
LOCUS      Homo sapiens BAC clone Rpl1-33P21 from 7, complete sequence.
AC072054      AC072054      GI:18056706
VERSION      HTG.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 154037)
AUTHORS      Sulston, J.E. and Wilson, R.
TITLE      Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      99063792
PUBMED      9847074
REFERENCE      2 (bases 1 to 154037)
AUTHORS      Du, H., Edwards, J., Haekenson, W. and Spalding, L.
TITLE      The sequence of Homo sapiens BAC clone Rpl1-33P21
JOURNAL      Unpublished (2001)
REFERENCE      3 (bases 1 to 154037)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (07-JUN-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE      4 (bases 1 to 154037)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (04-JUN-2002) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE      5 (bases 1 to 154037)
AUTHORS      Waterston, R.
TITLE      Direct Submission
JOURNAL      Submitted (30-JUN-2002) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE      6 (bases 1 to 154037)
AUTHORS      Waterston, R.
TITLE      Direct Submission
JOURNAL      Submitted (21-FEB-2002) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE      7 (bases 1 to 154037)
AUTHORS      Waterston, R.
TITLE      Direct Submission
JOURNAL      Submitted (26-APR-2003) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE      8 (bases 1 to 154037)
AUTHORS      Wilson, R.
TITLE      Direct Submission
JOURNAL      Submitted (08-OCT-2003) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Jan 4, 2002 this sequence version replaced gi:13786481.

COMMENT      ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu
              Contact: sapiens@wustl.wustl.edu
              ----- Summary Statistics
              -----
              Center project name: H_NH0033P21

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:
 The RPl1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateno, M., Caranese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.choi.org>

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RPl3-552B11 the clone sequenced to the right is RPl1-32P3, 2000 bp overlap. Actual start of this clone is at base position 22189 of RPl3-552B11 actual end is at base position 2000 of RPl1-32P3.

FEATURES

source

```

1.154037
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RPl1-33P21"
/clone_lib="RPl1-11"
1.550
/rpc_family="L1"
repeat_region
789..1317
/rpc_family="L1"
repeat_region
4289..4432
/rpc_family="BRVL"
repeat_region
4964..5214
/rpc_family="MIR"
repeat_region
5233..5483
/rpc_family="Alu"
repeat_region
5500..5520
/rpc_family="L1"
repeat_region
5994..6283
/rpc_family="AT-rich"
repeat_region
6337..6661
/rpc_family="Alu"
repeat_region
6991..7114
/rpc_family="L1"
repeat_region
7401..7634
/rpc_family="Alu"
repeat_region
7641..7759
/rpc_family="L2"
repeat_region
7797..7843
/rpc_family="L1"
repeat_region
7964..8144

```

```
repeat_region      /rpt_family="MALR"
                    8148..8458
                    /rpt_family="Alu"
repeat_region      8451..8747
                    /rpt_family="Alu"
repeat_region      8755..8806
                    /rpt_family="GGGA)n"
repeat_region      9349..9499
                    /rpt_family="L1"
repeat_region      9513..9762
                    /rpt_family="Alu"
repeat_region      9768..9936
                    /rpt_family="Alu"
repeat_region      9938..10243
                    /rpt_family="Alu"
repeat_region      10278..10576
                    /rpt_family="L1"
repeat_region      10577..10897
                    /rpt_family="Alu"
repeat_region      10898..11020
                    /rpt_family="L1"
repeat_region      11187..11411
                    /rpt_family="L1"
repeat_region      11412..11711
                    /rpt_family="Alu"
repeat_region      11712..11961
                    /rpt_family="L1"
repeat_region      11973..12264
                    /rpt_family="Alu"
repeat_region      12265..12610
                    /rpt_family="L1"
repeat_region      12611..12907
                    /rpt_family="Alu"
repeat_region      12908..13739
                    /rpt_family="L1"
repeat_region      13743..13876
                    /rpt_family="L1"
repeat_region      13885..13934
                    /rpt_family="MER1_type"
repeat_region      14077..14139
                    /rpt_family="MIR"
repeat_region      14576..14863
                    /rpt_family="Alu"
repeat_region      15079..15379
                    /rpt_family="Alu"
repeat_region      15648..16009
                    /rpt_family="Alu"
repeat_region      16015..16322
                    /rpt_family="Alu"
repeat_region      16326..16454
                    /rpt_family="Alu"
repeat_region      17256..17384
                    /rpt_family="L1"
repeat_region      17389..17628
                    /rpt_family="Alu"
repeat_region      17639..18045
                    /rpt_family="ERV.L"
repeat_region      18046..18454
                    /rpt_family="TA)n"
repeat_region      18455..18515
                    /rpt_family="ERV.L"
repeat_region      18516..18576
                    /rpt_family="Alu"
repeat_region      18643..18952
                    /rpt_family="Alu"
repeat_region      19052..19447
                    /rpt_family="MALR"
```

```
Query Match      99.8%; Score 877.4; DB 9; Length 154037;
Best Local Similarity 99.9%; Pred.No. 4,1e-262;
Matches 878; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGTCTGTATGAAGAAGATGACAT 60

```
|||||
Db      47417 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGTCTGTATGAAGAAGATGACAT 47358
Qy      61 AGAGACTCCATTTTGTATGTAAGAAAATTTCTGCGCTTGATTCGTATATC 120
Db      47357 AGAGACTCCATTTTGTATGTAAGAAAATTTCTGCGCTTGATTCGTATATC 47298
Qy      121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGTCCACTCAGGTTAA 180
Db      47297 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGTCCACTCAGGTTAA 47238
Qy      181 ATGATTAAAGGCGGTGACAGATGTCTTTGTTAAAGATGCTTGAAGCAGATGCTC 240
Db      47237 ATGATTAAAGGCGGTGACAGATGTCTTTGTTAAAGATGCTTGAAGCAGATGCTC 47178
Qy      241 CTTAAGATCATACACATCCCTTAATCTCAAGTACCCAGGACACAAAACGTGGGAAG 300
Db      47177 CTTAAGATCATACACATCCCTTAATCTCAAGTACCCAGGACACAAAACGTGGGAAG 47118
Qy      301 CCGCAGGAGACTGTGCTAGGAAAGCAGGTAATGTCACAGTTTCTCCCATGTGATAG 360
Db      47117 CCGCAGGAGACTGTGCTAGGAAAGCAGGTAATGTCACAGTTTCTCCCATGTGATAG 47058
Qy      361 CTTGAATATGCTCTGCTGAGAAAGAAAGACTGACCGTCCCGACCCGACACCGGTA 420
Db      47057 CTTGAATATGCTCTGCTGAGAAAGAAAGACTGACCGTCCCGACCCGACACCGGTA 46998
Qy      421 AAGGCTCTGCTGAGAGATTTGTAAAGAGAGAAATGCTCTTGACGTGAACACA 480
Db      46997 AAGGCTCTGCTGAGAGATTTGTAAAGAGAGAAATGCTCTTGACGTGAACACA 46938
Qy      481 AGAGAAAGGCAATGTCCTGCTCCGCTGCTCCCGGCAATGGAATGTCGATTAAGACC 540
Db      46937 AGAGAAAGGCAATGTCCTGCTCCGCTGCTCCCGGCAATGGAATGTCGATTAAGACC 46878
Qy      541 GATTGTATGCTCATCTACTAGATAGAGAAAACCGCTTAGGCTGAGGTGGAGACT 600
Db      46877 GATTGTATGCTCATCTACTAGATAGAGAAAACCGCTTAGGCTGAGGTGGAGACT 46818
Qy      601 GCGGCGACCAATACTGCTTTGTAAAGCACTGAGATGTTATGTATGATATCTAAAG 660
Db      46817 GCGGCGACCAATACTGCTTTGTAAAGCACTGAGATGTTATGTATGATATCTAAAG 46758
Qy      661 CACAGCACTTAATGCTTTTACATGCTATGATGCAAGACCTTGTCACATGTTGCT 720
Db      46757 CACAGCACTTAATGCTTTTACATGCTATGATGCAAGACCTTGTCACATGTTGCT 46698
Qy      721 GCTGACCTCTCCCAACATTTGTTGACCCCTGACACATCCCTCTTGAGAAAACAC 780
Db      46697 GCTGACCTCTCCCAACATTTGTTGACCCCTGACACATCCCTCTTGAGAAAACAC 46638
Qy      781 CCACAGATGATCAATAAATCTAAGGAAGCTCAGAGCGTGGCGGATCTTCATATGCTG 840
Db      46637 CCACAGATGATCAATAAATCTAAGGAAGCTCAGAGCGTGGCGGATCTTCATATGCTG 46578
Qy      841 AAGCGTGTCCCGGGTCCCTCTTCTTCTCTATA 879
Db      46577 AAGCGTGTCCCGGGTCCCTCTTCTTCTCTATA 46539
```

RESULT 6
HENV17834
LOCUS HENV17834 8300 bp DNA linear VRL 21-OCT-1999
DEFINITION Human endogenous retrovirus K (HERV-K) elements, clone SD-C7-34LTR.
ACCESSION Y17834
VERSION Y17834.1 GI:4185945
KEYWORDS env gene; env protein; gag gene; gag protein; long terminal repeat;
SOURCE Human endogenous retrovirus K
ORGANISM Human endogenous retrovirus K
REFERENCE 1
AUTHORS Tonjes, R. R., Czauderna, F. and Kurth, R.

TITLE Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K
JOURNAL J. Virol. 73 (11), 9187-9195 (1999)
MEDLINE 9945825
PUBMED 10516026
REFERENCE 2 (bases 1 to 8300)
AUTHORS Toenjes R.R.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1998) R.R. Toenjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
COMMENT Related sequence Y10390.
FEATURES
source location/Y10390
1..8300
/organism="Human endogenous retrovirus K"
/proviral
/mol_type="genomic DNA"
/specific_host="human"
/db_xref="taxon:45617"
/clone="SD-C7-34LTR"
/note="assigned to human chromosome 7"
32..2032
/gene="gag"
32..2032
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="CAA76884.1"
/db_xref="GI:4185946"
/db_xref="GOA:Q9YNA6"
/db_xref="SPTREMBL:Q9YNA6"
/translation="MGCTKSKSKVSYSTFIKILKRGVNVSTNLIKLPQIIHQ
FCFMPFQGLDLKDMKRIKIGELKQGRKSNIIPLTVNWMALITKALBEPQIBEDSV
SVSDAPGSCIIDCENETRKKSQKTEELHCEYVAEPMAOSTQVNDVNOLOEVIYFET
LKLEGGKPELVGPSSESKRGTSPLPAQGVVTLQPKQVKNKTQVPAQVYAPL
QYRPPPSOYQYRGMPAPQGRAPYPPQPRRLNPLTPRPSQSKLHEIIDSKEGSD
TBAQPPVTLBMPRGSGAGSEPTTAAKYSPIIKLDMKEGVQVGPVNPYMT
LDSIAGHMLIPYDWEIILAKSSISPSQIFQFTWVLDGVOEYRRRRANPPVNDI
DQLLIGQNNSTISQALQNEALIEVRAICLRAMEKIQDPGSTCEFNTRGSKSE
YDPFVARLQVQAKSINDEKARKVIELMAYENANPECSQAIKPLKGVAGDVISE
YVACDGIIGAMHAKMALMAQIITGVVAGGVRTGRCYNGQGIHLKKNCPVLNKN
ITVQATTTGREPDLCPCKKGGKMAQSGSKPKNGQPSLGNQGRQGPQAPQQTAF
PIQPFVPGSGGQGPPLSYVFGISQLPQVNNCPQAAVQ"
3051..5669
/gene="pol"
3051..5669
/gene="pol"
/codon_start=1
/product="pol protein"
/protein_id="CAA76885.1"
/db_xref="GI:4185947"
/db_xref="GOA:Q9WJR4"
/db_xref="SPTREMBL:Q9WJR4"
/translation="MLTDLRAVNAVIOPMGPILOGLPSPAMIPKMDPLIIIDKCEFF
TTPABODCEKFAFTIPAINNKBPATPQKVLPGQMLNPTICQTFPGALDPVREK
PSDCYIHCIDDLICAAETDKLIDCTTFLQAEVANAAGLIAADKIOSTSPHYLQ
IENKIKIPQKIEIRKDLTKLNDQKLGIDINMIRPLTGLPTAMSLFSLIGSDSL
NSRMLTPBAITKEIKVBEKIQSAQINRIDLPLAPLQIIPATASPLGIIQVTDVSE
WSFLPHSTVFTFLYLDQIALILGOTRLRIKLGNDPDIKVLPTQEOVROAFINS
AMKILANFVGIIDNHYPKTKIPOFLKLTWMLPKIRREPLENALVTFDGSNGK
AYGPKERVKITPQSOARAEVAVITVLDDPQPINIISDSAVVVOATDVTALIK
YSMDQNLQNLQCTVRKRNPFYITTHBAITNLPGLPTKANEQDLVLSALIK
QELHALAHVNAAGLKNFVDVTKQAKDIOHCTQCVLHLPTQEAQNPGLCPNALM
QMDVTHVPSRGLSVYHVTVDYSHFIMACQGESISHKHLSCFAMGVPEKIK
TDNGPYCSAKFOKLSQMKISHTTGIPYNSGOALVERTNRLKTQLVQKSGDSE
ECTTPQMLALATYLANLINYNQTTGSAOHLTGKNSPHGKLIIMKDNKNKWE
IGVITWGRGACSGENOLPWIIPRHLKTYNEPIRDKSGTSATETTSQSTYDS
ODEQNGDVRRTDEVAIHQGBRANLGTTKADAVSYKISREHKSDTNPREYACSLDD
CINGKSPYACRSSCS"
gene 5386..7470
/gene="env"
5386..7470
/gene="env"
/codon_start=1
CDS
gene

/product="env protein"
/protein_id="CAA76886.1"
/db_xref="GI:4185948"
/db_xref="GI:4185948"
/db_xref="GOA:Q9YNA5"
/translation="MQRKAPRRRRHRNRNAPLTHKNQVTSSEOKLPTSKKAPPT
MAQKLTQLATYILENTYVOTQPSMILAAIMVMMVSLPMDAGAAANVYMYAV
PPEPLRAVYMDNPTVEYVNDVVPGLPDRCPAKPEBEGMINISIGYHPITL
GRAPGIMPVQNVMLVEVPVSPICRFTYHMGSLRPVNVLODFSYQSLKPRK
GXCPKEIPRESKNTLEVWECVANSAYILONNEFGTIDMAPRGQFYNSCGQDS
CPSAQS PAVDSDLTRESLDKHKHKLQSFYPMSEKGIISTPRKIVSPVSGEHEL
MRITVSHHRIWMSGNQTLTETDRKPFYITIDNSSLTVPLQSCVCKPYMVGNIIVK
PQSOTTCENCRLLCTIDSTFWQHRILVRAEGWIPVSMRDPREASVYITLEV
LKGVLNRSKRFITTLIAVTMGLIAVTAATAVAGVALHSSVQSVNPNDRKSTRILN
SSSIDQKLANQINDLRQTVIWMGDRLSLEHFRQJCCMNTSDFCITPQIYNESLH
MDVNRHLQGREBNTLIDISKLEQIFEASKALNLPQTEALVAGVADGIANLNPVTW
VKTIQSTIILNLIILVLCFLILLVCRCTQQLRRDSDBERAMWTVAIVSKRGVNG
KSKRDQIVTVSV"
7425..8300
/note="3' LTR"
ORIGIN
Query Match 99.7%; Score 876; DB 14; Length 8300;
Best Local Similarity 100.0%; Pred. No. 98-262;
Matches 876; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAGCAAGAGAGATCAGATTGTACTGTGTCTGTGTAGAAAAGTAGACAT 60
Db 7425 TGTGGGAAAAGCAAGAGAGATCAGATTGTACTGTGTCTGTGTAGAAAAGTAGACAT 7484
QY 61 AGNAGATCCTATTGTGTATGTACTAAGAAAATTTCTTCGCTTGAAGATTCTGTATC 120
Db 7485 AGNAGATCCTATTGTGTATGTACTAAGAAAATTTCTTCGCTTGAAGATTCTGTATC 7544
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACAATGTCGTGTCCACTCGAGGTTAA 180
Db 7545 TATGACCTTACCCCAACCCCGTCTCTGAAACAATGTCGTGTCCACTCGAGGTTAA 7604
QY 181 ATGATTAAGAGGGGGTGACAGATGTCTTTGTTAAACAGATGCTTGAAGCAGATGCTC 240
Db 7605 ATGATTAAGAGGGGGTGACAGATGTCTTTGTTAAACAGATGCTTGAAGCAGATGCTC 7664
QY 241 CTTAAGAGTCATACCACTCCCTAATCTCAAGTACCAGGGAACAAAACCTCGGAAAG 300
Db 7665 CTTAAGAGTCATACCACTCCCTAATCTCAAGTACCAGGGAACAAAACCTCGGAAAG 7724
QY 301 CCGGAGGACCTCTGCTAGAGAAAGCAGGTAATGTCCAAAGTTTCTCCCATGTGATAG 360
Db 7725 CCGGAGGACCTCTGCTAGAGAAAGCAGGTAATGTCCAAAGTTTCTCCCATGTGATAG 7784
QY 361 CCTGAATATAGGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCCGACCCGTA 420
Db 7785 CCTGAATATAGGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCCGACCCGTA 7844
QY 421 AAGGCTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGCAAGTTGAGACA 480
Db 7845 AAGGCTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGCAAGTTGAGACA 7904
QY 481 AGAGGAAGGACATGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 7905 AGAGGAAGGACATGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7964
QY 541 GATTGATGCTTCATCTAGATAGAGGAAACCGCTTAGAGGCTGAGAGGAGGAGCCT 600
Db 7965 GATTGATGCTTCATCTAGATAGAGGAAACCGCTTAGAGGCTGAGAGGAGGAGCCT 8024
QY 601 GCGGAGAGCAATCTGCTTTGTAAGACCTGAGATGTTTATGTGTATGCAATATCTAAAG 660
Db 8025 GCGGAGAGCAATCTGCTTTGTAAGACCTGAGATGTTTATGTGTATGCAATATCTAAAG 8084
QY 661 CACAGACTTAATCCTTTACATGCTATGATGCAAGACCTTTGTTACATGTTGTCT 720
Db 8085 CACAGACTTAATCCTTTACATGCTATGATGCAAGACCTTTGTTACATGTTGTCT 8144

Oy	721	GCTGACCCCTTCCCCACAATTGTCTTGAGACCGCTGACATGCCCTTTGGAAAAAC	780
Dd	8145	GCTAACCTCTTCCCACAATTGTCTTGAGACCGCTGACATGCCCTTTGGAAAAAC	82040
Oy	781	CCACAGATGATCAGTAATACTAAAGGAACCTCAGAGCGTCGGGATCTCCATATGCTG	840
Dd	8205	CCACAGATGATCAGTAATACTAAAGGAACCTCAGAGCGTCGGGATCTCCATATGCTG	82640
Oy	841	AACGCTGTTCCCGGGTCCCTCTTCTTCTTCTTCT	876
Dd	8265	AACGCTGTTCCCGGGTCCCTCTTCTTCTTCTTCT	8300
RESULT 7			
LOCUS	AF940464		
DEFINITION	Homo sapiens HERV-K' long terminal repeat, complete sequence; and envelope glycoprotein gene, partial cds.		
ACCESSION	AF940464		
VERSION	AF940464.1		
KEYWORDS	GI:20067082		
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Reus,K., Stuhrr,T., Mayer,J. and Meese,E.U.		
TITLE	Haploctype analysis of Human Endogenous Retroviruses: A genetic variant of HERV-K(HML-2.HOM) with an intact YXDD motif of reverse transcriptase Unpublished		
JOURNAL	2 (bases 1 to 8708)		
REFERENCE	Reus,K., Stuhrr,T., Mayer,J. and Meese,E.U.		
AUTHORS	Direct Submission		
TITLE	Submitted (06-MAR-2002) Humangenetik, Universitaet des Saarlandes, Kirtbergstrasse, Homburg 66421, Germany		
JOURNAL	Location/Qualifiers		
FEATURES			
SOURCE	1..8708		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="7"		
	/note="from Asian sample"		
	/note="endogenous virus: HERV-K'"		
	239..>8708		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="7"		
	/note="endogenous _virus: HERV-K'; from Asian sample"		
	239..1206		
LTR	1350..3350		
misc_feature	/note="nonfunctional gag protein due to mutation"		
	3153..4155		
misc_feature	/note="nonfunctional protease protein due to mutation"		
	4115..6986		
misc_feature	/note="nonfunctional polymerase protein due to mutation in gag"		
	6988..->8708		
CDS	/codon_start=1		
	/product="envelope glycoprotein"		
	/protein_id="AA09525.1"		
	/db_xref="GI:20067083"		
	/translation="MNSSEMRKAPRRRRHRNRNAPLTHCKNKRVTSSEOKLPSTKK AEPPTAQKLQTLATLTENTKYTOTIPSPMLALMTIVSMVSLPMPGAAAANTY YVAVPEPLPIRAVTMDNPTEPVVNDWVPGPIDRCPAKPEEEMMINISIGHY PYIAVPRPGCLMPAVDNLVEVTPVSDPICFTYHWSGMSLRVVNYLDIFSOSL KFRKGRCPKEIPEKSNETVLWEBCVANSAVILONNEBGTIIDAPGOFYNCS GOTSCSAOVSPAVSDLTESLDKHKKLQSYPWEGEKGISRPRTIVPSVSGP EHPEIMLTIVAHHIRIMSGNOTLETDRPFYTIDASSLTYPLOS CYRPRYLVG NIYKPGSOTTCHNCRLTITDSTFPMQHRIILVRAREGWATFSMDRPEASPSVH ILTEVLGNVRSRPIFTLIATVMGLIATVAAGVALHSVGSVFVNWDQKNS		

[illegible]

DEFINITION Human endogenous retrovirus K (HERV-K) elements, clone C7.
ACCESSION Y17832
VERSION Y17832.2 GI:4581240
KEYWORDS env gene; env protein; gag gene; gag protein; long terminal repeat;
SOURCE Human endogenous retrovirus K
ORGANISM Human endogenous retrovirus K
VIRUSES; Retrovirdae.
REFERENCE
1 Tonjes R.R., Crauderma, F. and Kurth, R.
TITLE Expression of full-length human endogenous retrovirus type K
JOURNAL U. Virol. 73 (11), 9187-9195 (1999)
MEDLINE 99445825
PUBMED 10516026
REFERENCE
2 Toenjes R.R.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1998) R.R. Toenjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
REMARK Revised by (3)
REFERENCE
3 (bases 1 to 10029)
TITLE Toenjes R.R.
JOURNAL Direct Submission
Submitted (24-MAR-1999) R.R. Toenjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
COMMENT On Apr 12, 1999 this sequence version replaced gi:4185937.
Related sequence Y10390.
FEATURES
source
location/Qualifiers
1..10029
/organism="Human endogenous retrovirus K"
/proviral
/mol_type="genomic DNA"
/specific_host="human"
/db_xref="taxon:45617"
/clone="C7"
/clone_1lb="Pl (Genome Systems, St. Louis, MO, USA)"
/note="Assigned to human chromosome 7"
1..83
/rpt_family="Alu"
243..426
/rpt_family="Alu"
427..1394
/note="5' LTR"
1397..1414
1538..3538
/gene="gag"
1538..3538
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="CAA76879.1"
/db_xref="GI:4185938"
/db_xref="GOA:Q9YNB0"
/db_xref="SPTREMBL:Q9YNB0"
/translation="MGQTKSKIKSKVSYLSFIKILKRGVNVSTKNLTKLFOILIEQ
FCWPFPEOGTLDLDMRKIKELKQARKENIIPLTVMNMAIIKALBFOEBSY
SVSDAPSGCIIDNCNTRKSKOKETBGLHCEYVAEYMASTONVDVNOLOEVIYPT
LKI.BKGKPELVGPESEKRGSTPLPAGOVVTLLOPOKOVENKTOPPAVAAQYAEI
QYRPPEQYQYGRGMPAPRAGRAPYRPPPTRRLLNPTAPRQSKSHEIIDKSKGSD
TEAMQFPVLEPMPGSGAQEGEPVTEAKYSFSIKLMDKEGVAVQVGPSPYMT
LIDSIAHGRLIPIYDWBILAKSSLSPOQLOFTWIDVQOVRNRRAANPVNIDA
DOLIGCONRSLISQALMONEAIEQVRAICLRAMEKIOPGSTCFSTFNRQSGSEP
YLPFVARLOVAVOKSIDBKARKVIELMAYENANPCOSATIPILKGVAGSDVISE
YVRAQSIGAMKAMMAQITGVITGGVTRGRCVNCQOIGHKKKCPILANON
ITTOATTGRRPPDLCPKCKKAKMAQCSKFDKNGQPLSGNBQKQPAPOQTGAF
PIQPFVQGGQGGQPPUSVVFQGISQLPQYNNCPQPAAVQO"
gene
4557..7175
/gene="pol"
4557..7175
/gene="pol"
/codon_start=1
/product="pol protein"

/protein_id="CAA76879.1"
/db_xref="GI:4185939"
/db_xref="GOA:Q9YNB0"
/db_xref="SPTREMBL:Q9YNB0"
/translation="MLTDLRVANAVIGVOPLOPGLSPSPAMPKQNPILIIIDLDKCFE
TPIABODEKFAFTIIPAINKKEPAPRPFQWKLPOGMINSPTICOPFVGAALQVPERK
FSDCYIHCIDIDILCAETKDKLIDCYTTLQASVANAAGLAIASDKTQTSPPHYLQMO
IENRKIKPQKIEIRKDTLNDFOQLDINNIRPLGIPTVAMSNLSFSLNGSDLS
NSKRMILPEATEIKELVBERIKQSAQINRIDLPLQLIFATNHPGIIIIQNTDVE
WSFLPSHTVTEFTLYVDQIATLIGQTRLLIILKCGNDPKIVPLTKEQVQAIFNG
AMKIGANFVGIIDNHYPKTKIPOFLKTLIIPKTRREPNALTFVFGSSNGKA
AYNGPERVITKTPYOSAORAEIYAVTVLOPEOPINIIISDSYVVQAFDVEITALIK
YSMDQDNLQNLILQOTVKNRPPTITTHIRATNUPGPIITRANQDILVSSALIK
QELHALTVHNAAGLKNKFDVTWKQADIVQCTQCVLHPYBAGNPRGLCPNALW
QMDVTVPSEGRSLVYHVDVTSHPITWATQGESSTHYKHLCSFAWGVPEKIK
TUNGPGYKSAFOKFLSOWKI SHITGI PYNSOQALVETNRITLKTQLVQKGGDSK
ECTTPOMQNLALTYLNFPLINORTTSAEHLTGKNSPHEGKLIIMNDSKNKTE
IGKVITWGRGPAFCPSGENDLPVMIPTRHLPYNEPIPAKSTSAETESQSTDS
QDRQNDVPRIDVVAIHQGRANLQTTZBADAVYSIKSREHKDGNPREYAAICSLDD
CINGKSPYACRSCS"
6892..8976
/gene="env"
6892..8976
/gene="env"
/gene="env"
/codon_start=1
/product="env protein"
/protein_id="CAA76880.1"
/db_xref="GI:4185940"
/db_xref="GOA:Q9YNA9"
/db_xref="SPTREMBL:Q9YNA9"
/translation="MQRKAPRRRRRRRRAAPLTHKKNQVTSSEQKLBSTKAEPT
MAQLKLTOLATKYLENTKVTQTPESMLLAALMIVSMVVSLEPMPAAGAAAVYMAV
PEPPLIRAVTWMNDPTEVVVNDVSWVVPIDRCAPKPEBEGWMIISIDYHAPICL
GRAPGLMAYQNVMLVEPVPSPICRFTHYVSGMLRPRVNVLOFSYSLKPRPK
GKCPKEIRKESKNTKEVLWEECVANAVILONNEGTTIIDMAPROGVPVSGEHOES
CQSAQVSPAVSDLTESI.DKHKKKQASVPWMEWGGTIIIDMAPROGVPVSGEHOES
WRLTVASSHRTVSGQNTLETDRKFPYITIDNLSITLVPLQSCVKKPYMLVGNITIK
PQSQTTTENCRLTCTIDSTFWQHRILLVRAEGVMIIPVSMRPEWASVVIILVEY
LKGVLNRSKRFITLIVIMGLIAYATAVAVAGVLAHSSVQSVNFVMDKNSRTIRLN
SGSSIDOKLANQINDRQTVIMWGDRLMSLEHFOLOCDNNTSDFCITPQIYNSESH
MDVVRHLQGRENDLTLSIKLEQI.FEASKALNLYPGEATAGVADGANTANPYTW
VKRTGSTTINILILVLCPLILVRCITQQLRRDSHRRAMNTAVLSKRGANGV
KSKRDIVTVSV"
8931..9898
/note="3' LTR"
9899..9903
/rpt_family="Alu"
ORIGIN
Query Match 99.5%; Score 874.2; DB 14; Length 10029;
Best Local Similarity 99.7%; Pred. No. 3.3e-261;
Matches 876; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAGCAAGAGATCATGTTGCTGCTGTGTATAGAAAAGATAGCAT 60
DB 427 TGTGGGAAAAGCAAGAGATCATGTTGCTGCTGTGTATAGAAAAGATAGCAT 486
QY 61 AGAGAGCTCATTTTGTATGACTAAGAAAATTTCTTGCTTGAGATTCTGTTATC 120
DB 487 AGAGAGCTCATTTTGTATGACTAAGAAAATTTCTTGCTTGAGATTCTGTTATC 546
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACAATGCTGTGTCTCACTAGAGGTTAA 180
DB 547 TATGACCTTACCCCAACCCCGTCTCTGAAACAATGCTGTGTCTCACTAGAGGTTAA 606
QY 181 ATGATTTAAGGGGGGTGCAGAGATGCTGTTTAAACAGATGCTTGAAGCAGATGCTC 240
DB 607 ATGATTTAAGGGGGGTGCAGAGATGCTGTTTAAACAGATGCTTGAAGCAGATGCTC 666
QY 241 CTTAAGAGTCATACCACTCCCTAATCTCAAGTACCAGGAGACACAAAACCTCGGAAAG 300
DB 667 CTTAAGAGTCATACCACTCCCTAATCTCAAGTACCAGGAGACACAAAACCTCGGAAAG 726

QY 301 CCGCAGGAGCCTTGCTAGAGAAAGCCAGGATTTGCCAGTTCTCCCATGTGATG 360
DB 727 CCGCAGGAGCCTTGCTAGAGAAAGCCAGGATTTGCCAGTTCTCCCATGTGATG 786
QY 361 CCGGAAATATGCGCTCTGAGGAAAGAAAGCCGACCGTCCCGACCCGACCCGTA 420
DB 787 CCGGAAATATGCGCTCTGAGGAAAGAAAGCCGACCGTCCCGACCCGACCCGTA 846
QY 421 AAGGCTGTGCTGAGGAGATTAAGAGAGAAAGAAATGCTTGGAGTTGAGACA 480
DB 847 AAGGCTGTGCTGAGGAGATTAAGAGAGAAAGAAATGCTTGGAGTTGAGACA 906
QY 481 AAGAGAGGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 907 AAGAGAGGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 966
QY 541 GATTGTATGCTCCATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 967 GATTGTATGCTCCATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
QY 601 GCGGGGAGCAATCTGCTTTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 1027 GCGGGGAGCAATCTGCTTTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
QY 661 CACAGACCTTAATCTTACATTTCTATGATGAGAGAGAGAGAGAGAGAGAGAG 720
DB 1087 CACAGACCTTAATCTTACATTTCTATGATGAGAGAGAGAGAGAGAGAGAGAG 1146
QY 721 GCTGACCTTCTCCGACAAATTTGTTGACCTGACATCCCTCTTTCAGAGAAAC 780
DB 1147 GCTGACCTTCTCCGACAAATTTGTTGACCTGACATCCCTCTTTCAGAGAAAC 1206
QY 781 CCAAGATGATCACTAATTAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 1207 CCAAGATGATCACTAATTAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
QY 841 AAGCTGTTCCCGGGTCCCTCTTCTTCTTCTCTATA 879
DB 1267 AAGCTGTTCCCGGGTCCCTCTTCTTCTTCTCTATA 1305

RESULT 9
AF261945 6206 bp DNA linear PRI 26-SEP-2000
LOCUS Homo sapiens human endogenous retrovirus K108, partial sequence.
DEFINITION AF261945
ACCESSION AF261945.1 GI:10304107
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 6206)
Turner,G., Barbulescu,M., Seaman,M.I., Kidd,K.K. and Lenz,J.
TITLE A genetic variant of HERV-K108 in humans with a YXDD Motif in
reverse transcriptase
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 6206)
Turner,G., Barbulescu,M., Seaman,M.I., Kidd,K.K. and Lenz,J.
REFERENCE Direct Submission
JOURNAL Submitted (28-APR-2000) Molecular Genetics, Albert Einstein College
of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
FEATURES
source
1..6206
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region
48..6206
/rpt_family="human endogenous retrovirus K108"
48..1015

ORIGIN
LTR
Query Match 99.1%; Score 871; DB 9; Length 6206;

Best Local Similarity 99.4%; Pred. No. 3,2e-260;
Matches 874; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAGCAGAGAGATGATGATTTGATCTGTCTGTGTAGAAAAGATGACAT 60
DB 48 TGTGGGAAAAGCAGAGAGATGATGATTTGATCTGTGTGTAGAAAAGATGACAT 107
QY 61 AAGAGATCCATTTTGTATGTACTAAGAAAATTTCTTGCCTTGAGATTTGTTATC 120
DB 108 AAGAGATCCATTTTGTATGTACTAAGAAAATTTCTTGCCTTGAGATTTGTTATC 167
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTCTGTCTGATCACTCAGGTTAA 180
DB 168 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTCTGTCTGATCACTCAGGTTAA 227
QY 181 ATGGAATTAAGGGGGGTGACAGATGTGCTTTGTTAAAGATGCTTGAAGGAGATGCTC 240
DB 228 ATGGAATTAAGGGGGGTGACAGATGTGCTTTGTTAAAGATGCTTGAAGGAGATGCTC 287
QY 241 CTTAAGAGTATGACACATCTCCCTAATCTCAAGTACCCAGGAGACAAAACCTGCGAAGG 300
DB 288 CTTAAGAGTATGACACATCTCCCTAATCTCAAGTACCCAGGAGACAAAACCTGCGAAGG 347
QY 301 CCGCAGGAGCCTTGCTCTAGAGAAAGCCAGGATTTGTCAAAGTTTCTCCCATGTGATG 360
DB 348 CCGCAGGAGCCTTGCTCTAGAGAAAGCCAGGATTTGTCAAAGTTTCTCCCATGTGATG 407
QY 361 CCGGAAATATGCGCTCTGAGGAAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 408 CCGGAAATATGCGCTCTGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 467
QY 421 AAGGCTGTGCTGAGAGAGATTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 468 AAGGCTGTGCTGAGAGAGATTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 527
QY 481 AAGAGAGGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 528 AAGAGAGGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 587
QY 541 GATTGTATGCTCCATCTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 588 GATTGTATGCTCCATCTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 647
QY 601 GCGGGGAGCAATCTGCTTTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 648 GCGGGGAGCAATCTGCTTTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 707
QY 661 CACAGACCTTAATCTTACATTTCTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 708 CACAGACCTTAATCTTACATTTCTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 767
QY 721 GCTGACCTTCTCCGACAAATTTGTTGACCTGACATCCCTCTTTCAGAGAAAC 780
DB 768 GCTGACCTTCTCCGACAAATTTGTTGACCTGACATCCCTCTTTCAGAGAAAC 827
QY 781 CCAAGATGATCACTAATTAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 828 CCAAGATGATCACTAATTAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 887
QY 841 AAGCTGTTCCCGGGTCCCTCTTCTTCTTCTCTATA 879
DB 888 AAGCTGTTCCCGGGTCCCTCTTCTTCTTCTCTATA 926

RESULT 10
HERVORF 1505 bp mRNA linear VRL 06-APR-1995
LOCUS Human endogenous retrovirus mRNA for central open reading frame.
DEFINITION X82271
ACCESSION X82271.1 GI:757869
VERSION X82271.1
KEYWORDS open reading frame.
SOURCE Human endogenous retrovirus
ORGANISM Human endogenous retrovirus

REFERENCE 1
AUTHORS Lower, R., Tonjes, R.R., Korbmacher, C., Kurth, R. and Lower, J.
TITLE Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K
JOURNAL J. Virol. 69 (1), 141-149 (1995)
MEDLINE 95074858
PUBMED 7983704
REFERENCE 2 (bases 1 to 1505)
AUTHORS Tonjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1994) R.R. Tonjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
FEATURES
SOURCE 1. 1505
/organism="Human endogenous retrovirus"
/mol_type="mRNA"
/specific_host="Homo sapiens, patient GH, teratocarcinoma"
/db_xref="taxon:11827"
/note="HERV-K"
gene 1. 1505
/gene="unknown orf"
misc_feature 255..256
/gene="unknown orf"
/note="splice site"
CDS 273..590
/gene="unknown orf"
/protein_id="CAA57722.1"
/db_xref="GI:757870"
/db_xref="SPTREMBL:O69383"
/translation="MNPSEMQKAPRRRRHRNRPAPLTHKNNKNTVSEQMKLPSTK
AEPPTMAQLKLTQLATKYLENTKVTQTPESMLAALMIVMSVAGVPSSEBTATIE
NGP"
misc_feature 436
/gene="unknown orf"
/note="additional nucleotide"
misc_feature 512
/gene="unknown orf"
/note="additional nucleotide"
misc_feature 533..534
/gene="unknown orf"
/note="splice site"
ORIGIN
Query Match 99.0%; Score 870; DB 14; Length 1505;
Best Local Similarity 99.4%; Pred. No. 5.9e-260;
Matches 873; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAGTAGACAT 60
DB 628 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAGTAGACAT 687
QY 61 AGGAGATCTCATTTTGTATGTACTTAAGAAAATTCTTGCCTTGAGATTCTGTATC 120
DB 688 AGGAGATCTCATTTTGTATGTACTTAAGAAAATTCTTGCCTTGAGATTCTGTATC 747
QY 121 TATGACCTTACCCCAACCCCGTCTCTGGAACATGTGTGTCCACTCAGAGTTAA 180
DB 748 TATGACCTTACCCCAACCCCGTCTCTGGAACATGTGTGTCCACTCAGAGTTAA 807
QY 181 ATGAGTTAAGGCGGTGACAGATGTGCTTTGTAAAGATGCTTGAAGGACGATCTC 240
DB 808 ATGAGTTAAGGCGGTGACAGATGTGCTTTGTAAAGATGCTTGAAGGACGATCTC 867
QY 241 CTTAAGAGTCATCAACAATCTCTTATCTCAAGTACCAAGGACCAAAAATCTGGGAAG 300
DB 868 CTTAAGAGTCATCAACAATCTCTTATCTCAAGTACCAAGGACCAAAAATCTGGGAAG 927
QY 301 CCGGAGGACCTCTGCTTGAAGAAAGCAAGTATGTCACAGTTTCTCCCATGTGATAG 360
DB 928 CCGGAGGACCTCTGCTTGAAGAAAGCAAGTATGTCACAGTTTCTCCCATGTGATAG 987

QY 361 CTTGAATATGCTCTGTGGGAAAGGAAAGACTGACCGTCCCCAGCCGACACCCGTA 420
DB 988 TCTGAATATATGCTCTGTGGGAAAGGAAAGACTGACCGTCCCCAGCCGACACCCGTA 1047
QY 421 AAGGATCTGTGCTGAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTGAGTTGAGACA 480
DB 1048 AAGGATCTGTGCTGAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTGAGTTGAGACA 1107
QY 481 AGAGAAAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 1108 AGAGAAAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1167
QY 541 GATTGTATGCTCCATCTACTGAGATAGGAAAAACCGCTTAGGCGTGAAGTGGACCT 600
DB 1168 GATTGTATGCTCCATCTACTGAGATAGGAAAAACCGCTTAGGCGTGAAGTGGACCT 1227
QY 601 GCGGAGAGCAATCTGCTTTGTAAGCACTGAGATGTTATGTATGCTATCTAAAG 660
DB 1228 GCGGAGAGCAATCTGCTTTGTAAGCACTGAGATGTTATGTATGCTATCTAAAG 1287
QY 661 CACAGACTTAATCTTTACATGTCTATGATGATGATGATGATGATGATGATGATGAT 720
DB 1288 CACAGACTTAATCTTTACATGTCTATGATGATGATGATGATGATGATGATGATGAT 1347
QY 721 GCTGACCTCTCCCACAATTTGTTGTGACCTGACACATCCCTCTTGAGAAACAC 780
DB 1348 GCTGACCTCTCCCACAATTTGTTGTGACCTGACACATCCCTCTTGAGAAACAC 1407
QY 781 CCAAGATATGATGATTAATTAAGGAACTGAGAGCTGAGGAGATCCCTCATATGCTG 840
DB 1408 CCAAGATATGATTAATTAAGGAACTGAGAGCTGAGGAGATCCCTCATATGCTG 1467
QY 841 AAGCTGTTCCCGGAGTCCCTCTTCTTCTCTAT 878
DB 1468 AAGCTGTTCCCGGAGTCCCTCTTCTTCTCTAT 1505
RESULT 11
HERVENV 3366 bp mRNA linear VRL 30-MAR-1995
LOCUS Human endogenous retrovirus env mRNA.
DEFINITION X82272.1 GI:757871
ACCESSION X82272.1 GI:757871
VERSION
KEYWORDS env gene.
SOURCE Human endogenous retrovirus
ORGANISM Human endogenous retrovirus
REFERENCE 1
AUTHORS Lower, R., Tonjes, R.R., Korbmacher, C., Kurth, R. and Lower, J.
TITLE Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K
JOURNAL J. Virol. 69 (1), 141-149 (1995)
MEDLINE 95074858
PUBMED 7983704
REFERENCE 2 (bases 1 to 3366)
AUTHORS Tonjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1994) R.R. Tonjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
FEATURES
SOURCE 1. 3366
/organism="Human endogenous retrovirus"
/mol_type="mRNA"
/specific_host="Homo sapiens, patient GH, teratocarcinoma"
/db_xref="taxon:11827"
/note="HERV-K"
gene 1. 3366
/gene="env"
/db_xref="taxon:11827"
CDS 437..2536
/gene="env"
/codon_start=1
/protein_id="CAA57723.1"
/db_xref="GI:757872"

```
/db_xref="GOA:Q69384"
/translation="MNPSEMRKAPRRRRHRNRP.LTHKNKNVTSSEBOKL.PSTKK
ABPTPAQOLKLTOLATKYLENTKYTOTPESMLAALMYSMTVLSLMPAPAAANT
YMAVPPPLIRAVTMDNPLEVTVNDVWVPGPIDRCPKPEEBEGMINISGYH
PPLICGAPGLMAYONMVEPTVSPICFTYHMYSGMSLRVRVYLDQFYSORSL
KFRKGRPEKEIPEKSKTEVLWEEVANSVILONNEGCTIDAPROQFYNBS
GOTOSCSAQSYPVSDLTSLDKHKHKLQSYFPMWEGKISTPRPKVSPGSP
EHPMLRLTVASHHRIKISMGNOTLETDRKFPTIDINSLTVLSCVAPPMVLVG
NIVKPSQRTICENCRLTIDSTPNQRHILVVRBEGVITVSDRPREASPSYH
ILTEVLGLNRSKR1FTLAVIMGLAVATAVAVGALHSSVQSVMFVNDQKNS
TRLWNSQSSIDQKLANQINDLRQTVIMWGDLSLEHRFOLQCDMNTVDFCTEYQILN
ESEHMDMVRHLQGRDNLTLDISKLEQIFEASKAHNLVPGTEAIAGVADLANL
NPVTWKTIGSTTIINLILIVCLFCLLVCRCTQOLRRDSDHERAMMTVAIVSKK
GNVYKSRQDIIVTVSV"
```

misc_feature
600
/gene="env"
/note="additional nucleotide"
676
/gene="env"
/note="additional nucleotide"

misc_feature
676
/gene="env"
/note="additional nucleotide"

ORIGIN

Query Match 98.6%; Score 866.4; DB 14; Length 3366;
Best Local Similarity 99.3%; Pred. No. 8.4e-259;
Matches 870; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGTGGGAAAAGCAGAGATCATGATTGTTCTGTCTGTAGAAAAGATGACAT 60
DB 2491 TGTGGGAAAAGCAGAGATCATGATTGTTCTGTCTGTAGAAAAGATGACAT 2550
QY 61 AGGAGCTCCATTTGTATGATCAAGAAAATTTCTTCGCTGAGATCTGTATATC 120
DB 2551 AGGAGCTCCATTTGTATGATCAAGAAAATTTCTTCGCTGAGATCTGTATATC 2610
QY 121 TATGACCTTACCCCAACCCCGTCTCTGTGAAACATGCTGTGTCCACTCAGGCTTAA 180
DB 2611 TATGACCTTACCCCAACCCCGTCTCTGTGAAACATGCTGTGTCCACTCAGGCTTAA 2670
QY 181 ATGATTAAGGCGGTGAGATGTTCTTTTAAACAGATGCTTGAAGGACATGCTC 240
DB 2671 ATGATTAAGGCGGTGAGATGTTCTTTTAAACAGATGCTTGAAGGACATGCTC 2730
QY 241 CTTAAGAGTATCACTCCCTATCTCACTACCCAGGACACAAAATTGGGGAAG 300
DB 2731 CTTAAGAGTATCACTCCCTATCTCACTACCCAGGACACAAAATTGGGGAAG 2790
QY 301 CCGCAGGAGACTCTGCTTGAAGAACCCAGTATTGTCCAAAGTTTCTCCCATGTGATTG 360
DB 2791 CCGCAGGAGACTCTGCTTGAAGAACCCAGTATTGTTCMAAGTTTCTCCCATGTGAAAG 2850
QY 361 CTTGAAATATGCTCTGTGGGAGGAAAAGACTGACCGTCCCGACCCGACACCCGTA 420
DB 2851 CTTGAAATATGCTCTGTGGGAGGAAAAGACTGACCGTCCCGACCCGACACCCGTA 2910
QY 421 AAGGCTGTGCTGAGAGATTTAGTAAAGAGAAAGAAATGCTCTTGGCAGTTGAACA 480
DB 2911 AAGGCTGTGCTGAGAGATTTAGTAAAGAGAAAGAAATGCTCTTGGCAGTTGAACA 2970
QY 481 AGAGAGAGGAGCATCTGTCTCTGCTGCTCCCTGGGCAATGAAATGCTCGATATAAACCC 540
DB 2971 AGAGAGAGGAGCATCTGTCTCTGCTGCTCCCTGGGCAATGAAATGCTCGATATAAACCC 3030
QY 541 GATTGTATGCTCATCTAGATAGGAAAAACCCGCTTAGGAGCTGGAGTGGAGACT 600
DB 3031 GATTGTATGCTCATCTAGATAGGAAAAACCCGCTTAGGAGCTGGAGTGGAGACT 3090
QY 601 GCGGGCAGCATATCTGCTTTTAAAGCAGTGAATGTTATGTGTATGCAATCTTAAAG 660
DB 3091 GCGGGCAGCATATCTGCTTTTAAAGCAGTGAATGTTATGTGTATGCAATCTTAAAG 3150
QY 661 CACGACCTTAATCTTATCATGTTGATGCAAAAGACTTTGTTCACATGTTTGTCT 720

Db 3151 CACGACCTTAATCTTATCATGTTGATGCAAAAGACTTTGTTCACATGTTTGTCT 3210
QY 721 GCTACCTCTCCCAACAAATGCTGTGAGACCCGACACATCCCTCTTGAGAAACAC 780
Db 3211 GCTACCTCTCCCAACAAATGCTGTGAGACCCGACACATCCCTCTTGAGAAACAC 3270
QY 781 CCACAGATGATCACTAATTAATCTAAGGAACCTCAGAGCTGGCGGATCTTCATATGCTG 840
Db 3271 CCACAGATGATCACTAATTAATCTAAGGAACCTCAGAGCTGGCGGATCTTCATATGCTG 3330
QY 841 AACCTGATTCCTCCGCGGTCCCTCTTCTTCTCTCT 876
Db 3331 AACCTGATTCCTCCGCGGTCCCTCTTCTTCTCTCT 3366

RESULT 12
AC025420
LOCUS
DEFINITION
Human BAC library) complete sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AC025420 94145 bp DNA linear PRI 21-FEB-2003
Homo sapiens 12 BAC Rpl1-491C17 (Rosewell Park Cancer Institute
Human BAC library) complete sequence.
AC025420
AC025420.26 GI:13173480
HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 94145)
Mazny,D.M., Adams,C., Adio-Oduola,B., Ali-oshman,F.R., Allen,C.,
Alsbrooke,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarya,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonini,D., Bouck,J.,
Bowle,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chau,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Doulatwalie,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guayara,M.,
Guararane,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
Hodgeson,A., Hognes,M., Holloway,C., Hollins,B., Homs,F.,
Howard,S., Huber,J., Hulys,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Loulleged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Marondel,I., Martin,R.,
Martindale,A., Martinez,E., Massey,E., Mawliny,E., McLeod,M.P.,
Meador,M., Mei,G., Merscher,S., Metker,M., Miller,A., Miner,G.,
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,B., Nwokwenkwo,S.,
Ogulu,M., Okwono,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoubouk,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
Shoentari,N., Sisson,I., Sodergren,E., Sonake,T., Sparks,A.,
Stanley,H., Stone,H., Sutton,A., Svaltek,A., Tabor,P., Tamerisa,A.,
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Wellington,S., Williams,G., Williamson,A.,
Wieczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorilla,S., Zuercher,R., Zuercher,R., and Giddis,R.

TITLE
JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 94145)
Worley,K.C.

TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Mar 1, 2001 this sequence version replaced gi:12656668.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 94145
Phrap values in estimate: 93963
Average error rate (BCM-Phrap estimate): 4.17304e-05

Fraction of Phrap values less than 40 : 0.0184541
Number of consensus changing edits: 36
Number of N's in consensus : 0

Position	Original+Context	Edited+Context
5	cttcaaca (n) gnnnnnaa	cttcaaca (a) gacgtcgaa
7	ctnaacaang (n) ngcnnnaat	ctnaacaang (a) gctgcgaat
8	caacaangn (n) gnnnaaata	caacaangn (a) gctgcgaat
11	acaangnnc (n) nnaatcacg	acaangnnc (a) gcaatcacg
12	caangnncn (n) naataacag	caangnncn (a) caataacag
13	aangnncn (n) aaataacag	aaagacgct (c) aaataacag
33	ttctcttta (n) tnttctctc	ttctcttta (a) ttatctctc
36	ctttttat (n) ttctctctc	ctttctctc (a) ttatctctc
47	ttctctctc (n) cagttctcg	ttctctctc (c) cagttctcg
2176	ctaacatct (n) cagttctcg	ctaacatct (c) cagttctcg
7320	ttttgagaa (n) agctctgtc	ttttgagaa (g) agctctgtc
7415	aattcttag (n) ggcacattt	aattcttag (t) ggcacattt
16116	ttagctcca (n) cctctgcct	ttagctcca (g) cctctgcct
16128	cttcgcctc (n) tgcctctcc	cttcgcctc (t) tgcctctcc
16139	tgctctctc (n) cagggnnnn	tgctctctc (c) cagggnnnn
16145	ctccncaagg (n) nnnnnnnn	ctccncaagg (t) tctccncaag
16146	tcncaagg (n) nnnnnnnn	tcncaagg (t) tctccncaag
16147	ccncaagg (n) nnnnnnnn	ccncaagg (t) tctccncaag
16149	ncaggnnnc (n) nnnnnnnn	ncaggnnnc (t) tctccncaag
16150	caggnnncn (n) cmtgagaa	caggnnncn (c) cactgagaa
16152	ggnnnnncn (n) ntgagaaag	ggnnnnncn (a) ctgagaaag
43501	ggnnnnncn (n) ntgagaaag	ggnnnnncn (a) ctgagaaag
43501	ggnnnnncn (n) ntgagaaag	ggnnnnncn (a) ctgagaaag
48685	tggaataag (a) gaggtagata	tggaataag (g) gaggtagata
48685	atgagaagg (n) gaggtagata	atgagaagg (t) gaggtagata
64342	agtaattca (n) gttttcttt	agtaattca (t) gttttcttt
64346	atccangtc (n) ncttttnaa	atccangtc (t) ncttttnaa
64347	ttctnngtn (n) ctttttnaa	ttctnngtn (a) ctttttnaa
64356	gtttttttt (n) aannnncca	gtttttttt (t) aannnncca
64356	nncttttaa (n) gnnnnccat	nncttttaa (t) gnnnnccat
64359	ttttnaagc (n) nccatagtg	ttttnaagc (c) nccatagtg
64360	tttnaangn (n) ccataagtg	tttnaangn (c) ccataagtg
64381	ctccataag (n) ccataagtg	ctccataag (g) ccataagtg
67952	gtagaatta (n) ttctactat	gtagaatta (t) ttctactat
71109	agagagaata (n) cncactctg	agagagaata (a) cncactctg
71111	agcagaanc (n) cactgcac	agcagaanc (a) cactgcac
88461	ttgtatatac (n) gttatagag	ttgtatatac (a) gttatagag

----- Distribution of Quality < 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
1000	900	*							
800	800	*							
700	700	*							
600	600	*							
500	500	*							
400	400	*							
300	300	*							
200	200	*							
100	100	*							

Phrap Value Range

FEATURES

source

Version: 1.01 gxf.
Location:Qualifiers
1. .94145
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"

Query Match 98.4%; Score 864.6; DB 9; Length 94145;

Best Local Similarity 99.0%; Pred. No. 4e-258;
Matches 870; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

Oy 1 TGTGGGAAAAAGCAAGAGATCATGATTTGTAATCTGTCTGTATGAAAGAAAGATAGCAT 60
Db 45648 TGTGGGAAAAAGCAAGAGATCATGATTTGTAATCTGTCTGTATGAAAGAAAGATAGCAT 45707
Oy 61 AGGAGACTCCATTTGTTATGTAAGAAAAATTTCTCTCTGAGATCTGTGTAATC 120
Db 45708 AGGAGACTCCATTTGTTATGTAAGAAAAATTTCTCTCTGAGATCTGTGTAATC 45767
Oy 121 TATGACCTTACCCCAACCCCGTCTCTGAAAATGTGTCTGTCTGCACTCAGGTTAA 180
Db 45768 TATGACCTTACCCCAACCCCGTCTCTGAAAATGTGTCTGTCTGCACTCAGGTTAA 45827
Oy 181 ATGGATTAAAGGCGGTGAGATGTCTTTTAAACAGATGCTTGAAGAGACATGCTC 240
Db 45828 ATGGATTAAAGGCGGTGAGATGTCTTTTAAACAGATGCTTGAAGAGACATGCTC 45887
Oy 241 CTTAAGAGTCATCAACATCCCTAATCTCAAGTACCCAGGAGCAAAAACTGGGGAAG 300
Db 45888 CTTAAGAGTCATCAACATCCCTAATCTCAAGTACCCAGGAGCAAAAACTGGGGAAG 45947
Oy 301 CCGGAGGACCTCTGCTAGGAAAGCCAGATTTGTCCACGTTTCCCATGTAATAG 360
Db 45948 CCGGAGGACCTCTGCTAGGAAAGCCAGATTTGTCCACGTTTCCCATGTAATAG 46007
Oy 361 CTTGAATATGCTCTGCTGAGAAAGGAAAGCTGACCTGCTCCCAAGCCGACACCCGTA 420
Db 46008 TCTGAATATGCTCTGCTGAGAAAGGAAAGCTGACCTGCTCCCAAGCCGACACCCGTA 46067
Oy 421 AAGGCTGTGCTGAGAGAGATTAAGTAAAGAGAGAGATGCTCTTGCAGTTGAGACA 480
Db 46068 AAGGCTGTGCTGAGAGAGATTAAGTAAAGAGAGAGATGCTCTTGCAGTTGAGACA 46127
Oy 481 AGAGAAAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 46128 AGAGAAAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 46187
Oy 541 GATTGTATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 46188 GATTGTATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 46247
Oy 601 GCGGCGAGCATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 46248 GCGGCGAGCATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 46307
Oy 661 CACAGCACTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTT 720
Db 46308 CACAGCACTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTT 46367
Oy 721 GCTGACCTCTCTCCCAACATTTGTTGTGACCTGACATCCCTCTTCTGAGAAACAC 780
Db 46368 GCTGACCTCTCTCCCAACATTTGTTGTGACCTGACATCCCTCTTCTGAGAAACAC 46427
Oy 781 CCAAGATGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
Db 46428 CCAAGATGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 46487
Oy 841 AACGCTGTTCCCGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 879
Db 46488 AACGCTGTTCCCGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 46526

```

RESULT 13
AC110927 181916 bp DNA linear HTG 20-FEB-2002
LOCUS Homo sapiens chromosome UNK clone RP11-599B11, WORKING DRAFT
DEFINITION
SEQUENCE, 14 unordered pieces.
AC110927
AC110927.2 GI:18767601
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 181916)
Waterston, R.H.
The sequence of Homo sapiens clone

JOURNAL

2 (bases 1 to 181916)
Waterston, R.H.
Direct Submission

REFERENCE

Submitted (16-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

JOURNAL

3 (bases 1 to 181916)
Waterston, R.H.
Direct Submission

REFERENCE

Submitted (20-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

On Feb 20, 2002 this sequence version replaced gi:18693568.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H_NH0599B11
----- Summary Statistics -----
Sequencing vector: Plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 174138 bases at least Q40
Consensus quality: 176242 bases at least Q30
Consensus quality: 177435 bases at least Q20
Insert size: 18700; agarose-fp
Insert size: 180616; sum-of-contigs
Quality coverage: 0.00 in Q20 bases; sum-of-contigs
Quality coverage: 6.86 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1038: contig of 1038 bp in length
* 1039 1138: gap of unknown length
* 1139 2442: contig of 1304 bp in length
* 2443 2542: gap of unknown length
* 2543 4167: contig of 1625 bp in length
* 4168 4267: gap of unknown length
* 4268 5501: contig of 1234 bp in length
* 5502 5601: gap of unknown length
* 5602 9767: contig of 4166 bp in length
* 9768 9867: gap of unknown length
* 9868 16625: contig of 6758 bp in length
* 16626 16726: gap of unknown length
* 16726 21169: contig of 4444 bp in length
* 21170 21269: gap of unknown length
* 21270 27682: contig of 6413 bp in length
* 27683 27782: gap of unknown length
* 27783 38632: contig of 10850 bp in length
* 38633 38732: gap of unknown length
* 38733 54115: contig of 15383 bp in length
* 54116 66745: gap of unknown length
* 66745 66745: contig of 12530 bp in length
* 66746 66845: gap of unknown length

AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (29-APR-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 205317)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Feb 3, 2002 this sequence version replaced gi15624987.

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: sapiens@wustl.wustl.edu

Summary Statistics
Center project name: H_NH0518112
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu>

SOURCE INFORMATION:
The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osogawa, R., Woon, P. Y., Zhao, B., Frenken, E., Taten, M., Catanesse, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genome 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6.

FEATURES
source

1..205317
/location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-518112"
/clone_1fb="RPC1-11"
3228..3365
/rpt_family="MER1_type"
3307..3391
/rpt_family="MER1_type"
3413..3520
/rpt_family="MER1_type"
3522..3552
/rpt_family="MER1_type"
3590..3727
/rpt_family="TGAG)n"
5024..5315
/rpt_family="(CCGG)n"
repeat_region
/rpt_family="Alu"

repeat_region 5380..5433
/rpt_family="Alu"
repeat_region 5694..5878
/rpt_family="MIR"
repeat_region 6081..6519
/rpt_family="(TGG)n"
repeat_region 6558..6715
/rpt_family="(TGG)n"
repeat_region 6756..6853
/rpt_family="(TGG)n"
repeat_region 6875..7059
/rpt_family="(TGG)n"
repeat_region 7086..7949
/rpt_family="(TGG)n"
repeat_region 14663..14965
/rpt_family="(TGG)n"
repeat_region 16138..16164
/rpt_family="Alu"
repeat_region 17042..17107
/rpt_family="(A)n"
repeat_region 20446..20624
/rpt_family="(TGG)n"
repeat_region 20706..21006
/rpt_family="C-rich"
repeat_region 21038..21220
/rpt_family="C-rich"
repeat_region 21116..21847
/rpt_family="(CA)n"
misc_feature
/note="CpG island (%GC=62.6, o/e=1.44, #CpGs=49)"
21285..21466
/rpt_family="C-rich"
repeat_region 21454..21627
/rpt_family="C-rich"
repeat_region 21655..21805
/rpt_family="(CCCA)n"
repeat_region 21810..21830
/rpt_family="(CA)n"
repeat_region 21831..22358
/rpt_family="MER2_type"
repeat_region 22359..22657
/rpt_family="BRV1"
repeat_region 27420..27515
/rpt_family="MER2_type"
repeat_region 27526..27642
/rpt_family="(TGG)n"
repeat_region 27644..27823
/rpt_family="(TGG)n"
repeat_region 27834..27986
/rpt_family="(TGG)n"
repeat_region 28006..28078
/rpt_family="(TGG)n"
repeat_region 28080..28119
/rpt_family="T1"
repeat_region 28139..28275
/rpt_family="(TATATG)n"
repeat_region 30121..30429
/rpt_family="(TGG)n"
repeat_region 31275..31433
/rpt_family="Alu"
repeat_region 31447..31553
/rpt_family="MIR"
repeat_region 31578..31700
/rpt_family="L2"
repeat_region 33788..33688
/rpt_family="Alu"
misc_feature
/note="CpG island (%GC=66.0, o/e=0.70, #CpGs=62)"
34814..34986
/rpt_family="C-rich"
repeat_region 35032..35199
/rpt_family="C-rich"
repeat_region 35215..35385
/rpt_family="C-rich"
repeat_region 35338..35510
/rpt_family="C-rich"
repeat_region 35434..35604
/rpt_family="CT-rich"

```
repeat_region 35530..35716 /rpt_family="C-rich"  
repeat_region 35764..35852 /rpt_family="CT-rich"  
repeat_region 35854..36057 /rpt_family="L1"  
repeat_region 36068..36102 /rpt_family="L1"  
repeat_region 36103..36678 /rpt_family="(TG)n"  
repeat_region 36674..37337 /rpt_family="L1"  
repeat_region 39683..39983 /rpt_family="L1"  
repeat_region 40473..40709 /rpt_family="Alu"  
repeat_region 40473..40709 /rpt_family="(CCCCAG)n"
```

Query Match	98.4%;	Score 864.6;	DB 9;	Length 205317;
Best Local Similarity	99.0%;	Pred. No. 4.2e-258;		
Matches 870;	Conservative	0;	Mismatches 9;	Indels 0;
				Gaps 0;

OY	I	1769333AAAAGCACAAGAGATCAAGTGTACCTGTCCTGTGTGAAGAAGATGACAT	80
Db	172614	TGTGGGAAAAAGCAGAAGAGATCAGATTGTACTGTGTCTGTGTGAAGAAAGATGACAT	1726733
OY	61	AGGAGACTCCATTTTGTATTGTACTTAAGAAAAATCTTGCGCTTAGAATCTCTGTTAACT	120
Db	172674	AGGAGACTCCATTTTGTATTGTACTTAAGAAAAATCTTGCGCTTAGAATCTCTGTTAACT	1727333
OY	121	TATGACCTTACCCCCAACCCCGTGCCTCTGTGAACAATGTGCTGTGCACCTCAGGGTTAA	180
Db	172734	TATGACCTTACCCCCAACCCCGTGCCTCTGTGAACAATGTGCTGTGCACCTCAGAGTTGA	1727933
OY	181	ATGATTAAAGGCGGTGCAGATGTGCTTTGTTAAACAGATGCTTGAAGCAGCATGTCTC	240
Db	172794	ATGATTAAAGGCGGTGCAGATGTGCTTTGTTAAACAGATGCTTGAAGCAGCATGTCTC	1728533
OY	241	CTTAAAGATCATCACCACTCCCTAATCTTAAGTACCAGAGACAAAACTGCGGAAG	300
Db	172854	CTTAAAGATCATCACCACTCCCTAATCTTAAGTACCAGAGACAAAACTGCGGAAG	1729133
OY	301	CCGACGGAGACCTCTGCTTAGAAGACAGATTAATGTGCCAACGTTTTCCCACATGATAG	360
Db	172914	CCGACGGAGACCTCTGCTTAGAAGACAGATTAATGTGCCAACGTTTTCCCACATGATAG	1729733
OY	361	CCTGAATAATGAGCTCTGTGTGAAGGAAAGACCTGACCGTCCCGACCCGACACCCGTA	420
Db	172974	TCTGAATAATGAGCTCTGTGTGAAGGAAAGACCTGACCGTCCCGACCCGACACCCGTA	1730333
OY	421	AAGGCTCTGTGTGAGAGAGATTAGTAAAAAGAGAGAAATGCTCTTGACATTGAGACA	480
Db	173034	AAGGCTCTGTGTGAGAGAGATTAGTAAAAAGAGAGAAATGCTCTTGACATTGAGACA	1730933
OY	481	AGAGAAAGGCATCTGTCTCTGCTGTCCCTGSGCAATGAATGTCTCGGTATPAAAAACC	540
Db	173094	AGAGAAAGGCATCTGTCTCTGCTGTCCCTGSGCAATGAATGTCTCGGTATPAAAAACC	1731533
OY	541	GATTGTATGCTCCACTCTACTGAGATPAGGAAAAACCGCCTTAGGGCTGAGAGGTGGACCT	600
Db	173154	GATTGTATGCTCCACTCTACTGAGATPAGGAAAAACCGCCTTAGGGCTGAGAGGTGGACCT	1732133
OY	601	GCGGCGACCAATATCTGCTTTGTAAAGCATTGAGATTTTATGTGTATGCTATCTPAAAA	660
Db	173214	GCGGCGACCAATATCTGCTTTGTAAAGCATTGAGATTTTATGTGTATGCTATCTPAAAA	1732733
OY	661	CACAGCACTTAATCTTTTACATTTGCTATGATGACAAAGACCTTTGTTCACATGTTGYCT	720
Db	173274	CACAGCACTTAATCTTTTACATTTGCTATGATGACAAAGACCTTTGTTCACATGTTGYCT	1733333
OY	721	GCTGACCTCTCTCCCACAATTTGTCTTGTGACCTTGACACATCCCTCTTTCGAAAAAC	780
Db	173334	GCTGACCTCTCTCCCACAATTTGTCTTGTGACCTTGACACATCCCTCTTTCGAAAAAC	1733933

QY	781	CCACGATGATCAGTAATTAAGGGAATCTAGGCGGATCTCCATATGCTG	840
Db	173394	CCACGATGATCAGTAATTAAGGGAATCTAGGCGGATCTCCATATGCTG	173453
QY	841	AACGCTGGTCCCGGGTCCCTTCTTCTCTAT	879
Db	173454	AACGCTGGTCCCGGGTCCCTTCTTCTCTAT	173492
RESULT	15		
AL139404			
LOCUS			
DEFINITION	Human DNA sequence from clone Rpl1-376D11 on chromosome 10,	linear	PR1 12-SEP-2001
ACCESSION	AL139404		
VERSION	AL139404.9		
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 122913)		
JOURNAL	Whitehead, S.		
COMMENT	Direct Submission Submitted (12-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Sep 13, 2001 this sequence version replaced from gl:14586503. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr10 Rpl1-376D11 is from the library RRC1-11.2 constructed by the group of Pletier de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBAC3.6 IMPORTANT: This sequence is not the entire insert of clone Rpl1-376D11. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone Rpl1-376D11 is at 1 in this sequence. The true left end of clone Rpl3-27G4 is at 120914 in this sequence. The true right end of clone Rpl3-16H11 is at 12556 in this sequence.		
FEATURES			
source	Location/Qualifiers		
	1..122913		
	/organism="Homo sapiens"		
	/mol_type="Genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="10"		
	/clone="Rpl1-376D11"		
	/clone_jib="RRC1-11.2"		
	/clone="Rpl1-376D11"		
	/complement(18308..18734)		
	/note="match: GSS: Em:AOI75361"		
	/complement(18360..18734)		
misc_feature			
misc_feature			

```
misc_feature      complément (18308. 18734)
                  /note="match: GSS: Em:AQ175361"
misc_feature      complément (18360. 18734)
```

misc_feature

THIS PAGE BLANK (USP),

Web site: <http://www.shgc.stanford.edu>
 Contact: (Jackson, Mark) mcdpax1.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
 Series: IRAC Plate: 48 Row: P Column: 5
 This clone has the following problem: retained intron.

FEATURES

source

1. 1698
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5314057"
 /issue_type="Brain, hypothalamus"
 /clone_lib="NIH MGC_96"
 /lab_host="DH10B"
 /note="Vector: pBluescript"

ORIGIN

Query Match 94.0%; Score 826.2; DB 11; Length 1698;
 Best Local Similarity 96.2%; Pred. No. 3e-241;

Matches 846; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

QY 1 TGTGGGAAAAGCAAGAGATCAGATGTTACTGTCTGTGTAGAAAGAGTACAT 60
DB 805 TGTGGGAAAAGCAAGAGATCAGATGTTACTGTCTGTGTAGAAAGAGTACAT 864
QY 61 AGGAGATCTCCATTTTGTATGATCTAAGAAAATTTCTTGCCTTGAGATTCTTATC 120
DB 865 AGGAGATCTCCATTTTGTATGATCTAAGAAAATTTCTTGCCTTGAGATTCTTATC 924
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGCTGCTGCACACAGGTTAA 180
DB 925 TATGACCTTACCCCAACCCCGTCTCTGAAACATGCTGCTGCACACAGGTTAA 984
QY 181 ATGATTAAGGCGCGTCAGAGATGCTTTGTTAAAGATGCTTGAAGCAGATCTC 240
DB 985 ATGATTAAGGCGCGTCAGAGATGCTTTGTTAAAGATGCTTGAAGCAGATCTC 1044
QY 241 CTTAAGAGTCATCACCACTCCCTTAATCTCAAGTACCAAGGACAAACCTCGGAAAG 300
DB 1045 CTTAAGAGTCATCACCACTCCCTTAATCTCAAGTACCAAGGACAAACCTCGGAAAG 1104
QY 301 CCGAGAGGACCTCTGCTAGGAAAGCAGGATGTCACAGTTCTCCCATGTGATG 360
DB 1105 CCGAGAGGACCTCTGCTAGGAAAGCAGGATGTCACAGTTCTCCCATGTGATG 1164
QY 361 CTTGAAATATGCTCTGAGGAAAGGAAAGCATGACCGTCCCGACCGACACCCGTA 420
DB 1165 TCTGAAATATGCTCTGAGGAAAGGAAAGCATGACCGTCCCGACCGACACCCGTA 1224
QY 421 AAGGCTCTGCTGAGAGAGATTGTAAGAAAGAGAGATGCTTCTTGACATTGAGACA 480
DB 1225 AAGGCTCTGCTGAGAGAGATTGTAAGAAAGAGAGATGCTTCTTGACATTGAGACA 1284
QY 481 AGAGGAAGGATCTGCTCCGCTGCTGCGGCAATGGAATGTCGGATTAACCC 540
DB 1285 AGAGGAAGGATCTGCTCCGCTGCTGCGGCAATGGAATGTCGGATTAACCC 1344
QY 541 GATTGTATGCTCCATCTAGAGATAGGAAAACCCGCTTAGGGCTGAGGTGGAGACT 600
DB 1345 GATTGTATGCTCCATCTAGAGATAGGAAAACCCGCTTAGGGCTGAGGTGGAGACT 1404
QY 601 GCGGCGACGAATACTGCTTTGTAAGCACTGAGATGTTATGTGTATGATTAAG 660
DB 1405 GCGGCGACGAATACTGCTTTGTAAGCACTGAGATGTTATGTGTATGATTAAG 1464
QY 661 CACAGCACTTAATCCTTATGATGCTATGATGAAGACCTTGTTCACATGTTTCT 720
DB 1465 CACAGCACTTAATCCTTATGATGCTATGATGAAGACCTTGTTCACATGTTTCT 1524

```

QY 721 GCTGACCCCTCTCCCAACAATGTTGTGTGACCTGACACATCCCTCTTGAGAAACAC 780
 DB 1525 GCTGACCCCTCTCCCAACAATGTTGTGTGACCTGACACATCCCTCTTGAGAAACAC 1584
 QY 781 CCACAGATGATCGTAAATTAAGGAACTCAGAGGCTGGGGGATTCCTCATATGCTG 840
 DB 1585 CCACAGATGATCGTAAATTAAGGAACTCAGAGGCTGGGGGATTCCTCATATGCTG 1644
 QY 841 AACGCTGTTCCCGGGATCCCTCTTCTTCTCTATA 879
 DB 1645 AACGCTGTTCCCGGGATCCCTCTTCTCTCTATA 1683

RESULT 2
 BU618928/c 706 bp mRNA linear EST 23-SEP-2002
 LOCUS BU618928
 DEFINITION UI-H-FH1-bf1-b-11-0-UI.81 NCI CGAP FH1 Homo sapiens cDNA clone
 UI-H-FH1-bf1-b-11-0-UI 3', mRNA sequence.
 ACCESSION BU618928.1 GI:23285143
 VERSION BU618928
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index

JOURNAL
 COMMENT Unpublished (1997)
 Contact: Robert Straubeberg, Ph.D.
 Email: CGAPBS-r@mail.nih.gov
 Tissue Procurement: James Martin
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@uiowa.edu
 The following repetitive elements were found in this cDNA
 sequence: 18-706, >LTR5LTR/Retroviral (matched complement)
 Seq primer: M13 FORWARD
 PolyA=Yes.

FEATURES
 source

1. 706
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-FH1-bf1-b-11-0-UI"
 /issue_type="Cell Line"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP FH1"
 /note="Organ: Chondrosarcoma; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP FH1 is a normalized cDNA library obtained from a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing an EcoR I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGAATCCGCG. The cell line was provided by Dr. James Martin from the University of Iowa.
 TAG TISSUE=Human Chondrosarcoma Cell Line CS8 - Grade 1
 Chondrosarcoma
 TAG_LIB=UI-H-FH1
 TAG_SEQ=AGAATCCGCGC"

ORIGIN

Query Match 75.0%; Score 659.2; DB 13; Length 706;
 Best Local Similarity 97.2%; Pred. No. 2.9e-190;
 Matches 670; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 127 CTTACCCCAACCCCGTCTCTGAAACATGCTGTGCTCACTCAGGGTTAATGAT 186
 |||||
 DB 706 CTTACCCCAACCCCGTCTCTGAAACATGCTGTGCTCACTCAGGGTTAATGAT 647
 QY 187 TAAGGGGGTGCAGAGATGCTTTGTTAAACAGATGCTTGAAGCAGATGCTCTTAAG 246
 |||||
 DB 646 TAAGGGGGTGCAGAGATGCTTTGTTAAACAGATGCTTGAAGCAGATGCTCTTAAG 587
 QY 247 AGCATACACACCTCCCTAATCTCAAGTACCAGGACACAAAATCCGGAAAGCCGACG 306
 |||||
 DB 586 AGCATACACACCTCCCTAATCTCAAGTACCAGGACACAAAATCCGGAAAGCCGACG 527
 QY 307 GGAACCTCTGCTAGAAAGCCAGATATGTCACAGTTCTCCCATGTGATAGCTGAA 366
 |||||
 DB 526 GGAACCTCTGCTAGAAAGCCAGATATGTCACAGTTCTCCCATGTGATAGCTGAA 467
 QY 367 ATATGCGCTGTGGGAAAGGAAAGACCTGACCGTCCCGACCCGACACCCGTAAGGGT 426
 |||||
 DB 466 ATATGCGCTGTGGGATGGGAAAGACCTGACCGTCCCGACCCGACACCCGTAAGGGT 407
 QY 427 CTGTGCTGAGAGAGATTAATGTTAAAGAAAGAAATGCTCTTGAGTTGAGACAGAGGA 486
 |||||
 DB 406 CTGTGCTGAGAGAGATTAATGTTAAAGAAAGAAATGCTCTTGAGTTGAGACAGAGGA 347
 QY 487 AGGCATGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 546
 |||||
 DB 346 AGGCATGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 287
 QY 547 ATGCTCCATCTACTAGATAGGAAAAACCGCTTAAAGGCTGAGGAGTGGACCTGCGGGC 606
 |||||
 DB 286 ATGCTCCATCTACTAGATAGGAAAAACCGCTTAAAGGCTGAGGAGTGGACCTGCGGGC 227
 QY 607 AGCAATCTGCTTTGTAAGCACTGAGATGTTATGTTATGCTATCTAATTAAGACACAGC 666
 |||||
 DB 226 AGCAATCTGCTTTGTAAGCACTGAGATGTTATGTTATGCTATCTAATTAAGACACAGC 167
 QY 667 ACTTAATCTTTTACATGCTGATGATGCAAAAGACCTTTGTTCAATGTTGCTGCTGAC 726
 |||||
 DB 166 ACTTAATCTTTTACATGCTGATGATGCAAAAGACCTTTGTTCAATGTTGCTGCTGAC 107
 QY 727 CCTCTCCCAACAATGCTTGTGTGACCTGACACATCCCTCTTGAAGAAACACCAAG 786
 |||||
 DB 106 CCTCTCCCAACAATGCTTGTGTGACCTGACACATCCCTCTTGAAGAAACACCAAG 47
 QY 787 ATGATCAGTAATTAATTAAGGAACTCAGA 815
 |||||
 DB 46 ATGATCAGTAATTAATTAAGGAACTCAGA 18

RESULT 3
 BM981862/c 707 bp mRNA linear EST 21-FEB-2003
 LOCUS
 DEFINITION
 UI-CF-EN1-adj-g-12-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
 BM981862
 ACCESSION
 UI-CF-EN1-adj-g-12-0-UI.3, mRNA sequence.
 VERSION
 BM981862.1 GI:19604781
 KEYWORDS
 EST.
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 707)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL
 Genome Res. 6 (9), 791-806 (1996)
 MEDLINE
 97044477
 PUBMED
 8889548

COMMENT
 Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 CDNA Distribution: Researchers may obtain clones from Research
 Genetics (www.research.com) or from Open Biosystems
 (www.openbiosystems.com).
 The following repetitive elements were found in this cDNA
 sequence: 17-707, >L1R5#LTR/Retroviral (matched complement)
 Seq primer: M13 FORWARD
 POLYA=yes.

FEATURES
 source
 location/Qualifiers
 1..707
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-EN1-adj-g-12-0-UI"
 /issue_type="Primary Lung Cystic Fibrosis Epithelial
 Cells"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-EN1"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: Ecor I; Site 2: Not I;
 UI-CF-EN1 is a normalized cDNA library containing the
 following tissue(s): Primary Lung Cystic Fibrosis
 Epithelial Cells. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an Ecor I adaptor, digested with Not
 I, and cloned directionally into pT73-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dri)18 tail. The
 sequence tag for this library is CTGCTCAGT.
 TAG-TISSUE=Human Lung Epithelial Cell Lines untreated LPS
 6hr to LPS 24h
 TAG_LIB=UI-CF-EN1
 TAG_SEQ=CTGCTCAGT"

ORIGIN
 Query Match 74.1%; Score 651.2; DB 12; Length 707;
 Best Local Similarity 97.3%; Pred. No. 8.2e-188;
 Matches 673; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 151 GAAACATGCTGCTGTCACCTCAGAGGTTAAATGATTAAGGCGCTGACAGATGCTTT 210
 |||||
 DB 707 GAAACATGCTGCTGTCACCTCAGAGGTTAAATGATTAAGGCGCTGACAGATGCTTT 649
 QY 211 GTTAAACAGATGCTTGAAGGACGATCTCTTAAAGTATCAACCACTCCCTAATCTCA 270
 |||||
 DB 648 GTTAAACAGATGCTTGAAGGACGATCTCTTAAAGTATCAACCACTCCCTAATCTCA 589
 QY 271 AGTACCAGGACACAAAACCTGCGAAGGCGCGAGGACCTCTGCTTGAAGAAAGCCAGG 330
 |||||
 DB 588 AGTACCAGGACACAAAACCTGCGAAGGCGCGAGGACCTCTGCTTGAAGAAAGCCAGG 529
 QY 331 TATTGTCCAAAGTTTCTCCCATGTGATAGCTGAATATGAGCTGCGGAGGAAAG 390
 |||||
 DB 528 TATTGTCCAAAGTTTCTCCCATGTGATAGCTGAATATGAGCTGCGGAGGAAAG 469
 QY 391 ACTGACGCTCCCGACCGGACACCCGTAAGGGTGTGCTGAGAGATTAAGTAAAA 450
 |||||
 DB 468 ACTGACGCTCCCGACCGGACACCCGTAAGGGTGTGCTGAGAGATTAAGTAAAA 409

OY	451	GAGGAGGAAATGCCCTTGGCATTTGAGACAAAGGAAAGGCATCTGTCCTGGCTGTCCC	510
Db	408	GAGGAGGAAATGCCCTTGGCATTTGAGACAAAGGAAAGGCATCTGTCCTGGCTGTCCC	349
OY	511	TGGGCAATGGAATGCTCGGATTAACCCGATTTGATCTCCATCTACTGAGATAGGA	570
Db	348	TGGGCAATGGAATGCTCCATATTAACCCGATTTGATCTCCATCTACTGAGATAGGA	289
OY	571	AAAAACCGCTTNAAGGCTGAGGCTGGGACCTGGGGGACACMAATCTCTTTGTAAGCACT	630
Db	288	AAAAACCGCTTNAAGGCTGAGGCTGGGACCTGGGGGACACMAATCTCTTTGTAAGCAAT	229
OY	631	GAGATGTTATGTATGATGATATATCTAAAGACACAGACTTAATCCCTTACATTTGATATG	690
Db	228	GAGATGTTATGTATGATGATATCTTAAGACACAGACTTAATCTTTACATTTGATATG	169
OY	691	ATGCMAAGACCTTTGTTTCACATGTTTGTCTGCTGACCCCTCTCCCAACAATTGTCCTGTGA	750
Db	168	ATGCMAAGACCTTTGTTTCACGTTTGTCTGCTGACCCCTCTCCCACTAATTGTCCTGTGA	109
OY	751	CCCTGACACATCCCTCTTTTGAGAGAAACCCACAGATATCATGTAATATCTAAGGGAAC	810
Db	108	CCCTGACACATCCCTCTTTTGAGAGAAACCCACAAATGATCAATTAATACTAAGGGAAC	49
OY	811	TCAGAGGCTGGGGGAGATCTCCATATGCTGAA	842
Db	48	TCAGAGGCTGGGGGAGATCTCCATATGCTGAA	17

RESULT 4	AL565978/c	975 bp	mRNA	linear	EST 31-MAY-200
LOCUS	AL565978				
DEFINITION	AL565978 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone				
ACCESSION	CS0DF016Y004.3-PRIME, mRNA sequence.				
VERSION	AL565978				
KEYWORDS	AL565978.2 GI:31289151				
SOURCE	EST.				
ORGANISM	Homo sapiens (human)				
REFERENCE	Homo sapiens				
AUTHORS	Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
JOURNAL	1. (bases 1 to 975)				
COMMENT	Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.				
	Full-length cDNA libraries and normalization				
	Unpublished (2001)				
	On Feb 16, 2001 this sequence version replaced gi:12917885.				
	Contact: Genoscope				
	Genoscope - Centre National de Sequencage				
	BP 191 91006 Evry cedex - France				
	Email: seque@genoscope.cns.fr Web : www.genoscope.cns.fr				
	Library was constructed by Life Technologies, a division of				
	Invitrogen. This sequence belongs to sequence cluster 6209.r For				
	more information about this cluster, see				
	http://www.genoscope.cns.fr/				
	cgt-bln/cluster.cgi?seq=CS0DF016BG02NP1&cluster=6209.r Contact :				
	Feng Liang Email : fliang@lifeech.com URL :				
	http://fulllength.invitrogen.com/InvitroGen Corporation 1600				
	Paradise Avenue Genoscope sequence ID : CS0DF016BG02NP1.				

```

FEATURES
    source
        Location/Qualifiers
            1..975
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CSODF0167M04"
                /tissue_type="FETAL BRAIN"
                /dev_stage="fetal"
                /clone_1ib="Homo sapiens FETAL BRAIN"
                /note="Organ: brain; Vector: pcwmvSPORT 6; 1st strand cDNA
                    was primed with a NotI-oligo(dT) primer. Five prime end
                    enriched, double-strand cDNA was digested with Not I and
                    cloned into the Not I and EcoRV sites of the pcwmvSPORT 6
                    vector. Library was not normalized."

```

Query Match	73.6%;	Score 646.8;	DB 9;	Length 975;
Best Local Similarity	82.6%;	Pred. NO. 2.1e-186;		
Matches 690;	Conservative 58;	Mismatches 82;	Indels 5;	Gaps 3;

Qy	3	TGGGAAAAGAGAGAGATCAGATTGTCTTCTGTCGTGTAGAAAAGATGACATG	62
Dp	849	TTKKAAAAGSAAKAKAKASAKTTGTATSTGKCTGTCTATAAAAGATGAKCATRK	790
Qy	63	GAGACTCCATTTTGTATGTACTAGAAAAATTCCTTGCTGTAGATCTGTATCTA	122
Dp	789	GAKCTSCATTTTXXTSTGTASTAATTAATAAATSTTCKCTTATATCTKTAAKCTA	730
Qy	123	TGACCTTACCCCCCAACCCCGTCTCTGAAACATGTCGTGTGTC-CACCTAGGGTTAAA	181
Dp	729	TKACCTTACCCCAACCCCTTTCCTCTCTAAACANNNTCTATTTCTAAACNCGGTTTAAA	670
Qy	182	TGATTTAAGGCGGTGCGAGATGCTTGTGTAAACAGATCTTGAAAGCAGATGCTCC	241
Dp	669	TGATTTAAGGTTTNTNCANANTNTNCTTTGTTAAACAATGCTTGAAAGCAGATCTCC	610
Qy	242	TTAAGATCATCACCATCTCCCTAATCTCAGTACCCGAGACACAAAACTGCGGAGGC	301
Dp	609	TTAAAAATCATCACCACTCCCTAATCTCAAGTACCCGAGGABADPAAAACTCGCGAAGGC	550
Qy	302	CGCAGGAGACCTCTGCTCAGGAAAGCCAGATTTGTCCAAAGTTTCTCCCACTGTATAGC	361
Dp	549	CGCAGTACCTCTCSCTAKAGAAACCGATTTTNNCCAAAGTTNNCCCCGSAATGATKT	490
Qy	362	CTGAATATATGCGCTCGTGGGAAAGGAAAGACTGACCGTCCCGCAGCCCGACACCCGTAA	421
Dp	489	CTRAAATATATGCGCTCATKGGAAKKAATAAAACCTGACCATCCCSGVSCCNACACCKTAA	430
Qy	422	AGGCTCTGTGCTGAGAGAGATTAGTAAAGAGAAAGATGCCCTTGCTCAATTGAGACA	481
Dp	429	AGKTTBTGTCTGAGAGAGATTAGTAAAGAGAAAGCATTCBTCTTGCAATTGAGACA	370
Qy	482	GAGGAAGGCATCTGTCTCTGCGCTGTCCTGTGGGCAATGSAATGTCTGGTATTAACCCG	541
Dp	369	GAGGAAGGCATCTATMACTGTCCCGTCTCTGGGCAATGSAATGTCTGGTATTAACCCG	310
Qy	542	ATTGTATGCTCCATCTACTGAGATAGGGAATAACCGCCTTAGGCTGAGGTGGGACTG	601
Dp	309	ATTGTATGCTCCATCTACTGAGATAGGGAATAACCGCCTTAGGCTGAGGTGGGACTG	250
Qy	602	CGGGCAGCAATPACGCTTTGTAAAGCACTGAGATGTTATGATGATGATATCTTAAAGC	661
Dp	249	CAGGCACCAATPACGCTTTGTAAAGCACTGAGATGTTATGATGATGATATCTTAAAGC	190
Qy	662	ACAGCACTTAATCCTTACATTTGTCATGATGCAAAAGCTTTGTTCCATGTTGTCTG	721
Dp	189	ACAGCACTTGAATCTTAACTTGTCTATGATGCAAAAGCTTTTTCACCTGTTGTCTG	130
Qy	722	CTGACCTCTTCCCAACAATTGTCTTGTGACCTGACACATCCCTCTTGAGAAACACC	781
Dp	129	CTGACCTCTTCCCAANNBTATGTCTTGT--ACATATACATCTCKSTBBBHAGATCAC	72
Qy	782	CACGATGATCAGTAATCTAAGGGAACCTCAGAGGCTGGGGGATCCTCCATAT	836
Dp	71	BNCGAATGATCATAAATACTAAGGGAACCTCAG--TGGGGCGAATCTCCATAT	19

RESULT	5
LOCUS	BM986989/c
DEFINITION	704 bp mRNA linear EST 17-JUN-2003 Uf-H-DIO-atc-h-05-o-Uf.s1 NCl CGAP_DIO Homo sapiens CDNA clone
ACCESSION	BM986989
VERSION	IMAGE5562196.3, mRNA sequence.
KEYWORDS	BM986989
SOURCE	BM986989.1 GI:19709378
ORGANISM	EST. Homo sapiens (human) Homo sapiens Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 704)
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCI-GAP <http://www.ncbi.nlm.nih.gov/ncigap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/INL at: <http://image.lnl.gov>
The following repetitive elements were found in this cDNA
sequence: 16-704, >LTR5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

Location/Qualifiers
1..704
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5862196"
/tissue_type="Lung Focal Fibrosis"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP D10"
/note="Organ: Lung; Vector: pRTT3-Pac (Pharmacia) with a
modified polylinker; Site 1: Ecor I; Site 2: Not I;
NCI CGAP D10 is a cDNA library containing the following
tissue(s): A pool of lung Focal Fibrosis. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an Ecor I
adaptor, digested with Not I, and cloned directionally
into pRTT3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
ATACGGCTC.
TAG_TISSUE=lung with fibrosis
TAG_LIB=UI-H-D10
TAG_SEQ=ATACGGCTC"

ORIGIN

Query Match 71.8%; Score 631.4; DB 12; Length 704;
Best Local similarity 96.1%; Pred. No. 9.5e-182;
Matches 658; Conservative 0; Mismatches 26; Indels 1; Gaps 1;
132 CCCCACCCCGTCTCTCTGAACATGCTGTGTC-CAGTCAGGGTTAAATGATTAAG 190
133 CCCCACCCCGTCTCTCTCTGAACATGCTGTGTC-CAGTCAGGGTTAAATGATTAAG 190
700 CCCCAACCCATGCTCTCTGAACATGCGCTGTGCAAACTCAGGGTTAAATGATTAAG 641
191 GGGCGTGAGATGTGCTTTGTAAACGATGCTTGAAGCGACATGCTCTTAAAGATC 250
640 GGGCGTGAGATGTGCTTTGTAAACGATGCTTGAAGCGACATGCTCTTAAAGATC 581
251 ATCAACCACTCCCTATCTCAAGTACCCAGGGACACAAACATGCGGAGAGCCGACGGAGC 310
580 ATCAACCACTCCCTATCTCAAGTACCCAGGGACACAAACATGCGGAGAGCCGACGGAGC 521
311 CTCTGCTAGAGAAAGCCAGTATTGTCCCAAGTTTCTCCCATGTGATAGCTTGAATAT 370
520 CTCTGCTAGAGAAAGCCAGTATTGTCCCAAGTTTCTCCCATGTGATAGCTTGAATAT 461
371 GGGCTGCGGAGAGGAAAGACCTGACCGTCCCGAGCCGACACCCCTTAAAGGTCTGT 430
460 GGGCTGCGGAGAGGAAAGACCTGACCGTCCCGAGCCGACACCCCTTAAAGGTCTGT 401
431 GCTGAGAGATTAGTAAAGAGAAAGATGCTCTTGAAGTGAACAAGAGGAGGC 490
|||||

Db 400 GCTGAGAGATTAGTAAAGAGAAAGAAATGCCCTTTCAGATTGACAAAGAGAGGC 341
Qy 491 ATCTGCTCTGCGCTGTCCTCGGCAATGGAATGTCGGTATATAAACCCGATTGTATGC 550
Db 340 ATCTGCTCTGCGCTGTCCTCGGCAATGGAATGTCGGTATATAAACCCGATTGTATGC 281
Qy 551 TCATCTACTGAGATAGGAAAAACCGCTTAAAGGCTGAGGTGGACCTGGCGGACGA 610
Db 280 TCATCTACTGAGATAGGAAAAACCGCTTAAAGGCTGAGGTGGACCTGGCGGACGA 221
Qy 611 ATATCTGTTGTAAAGACATGATGTTTATGTATGCAATATCTAAACACAGACCT 670
Db 220 ATATCTGTTGTAAAGACATGATGTTTATGTATGCAATATCTAAACACAGACCT 161
Qy 671 ATCTCTTACATGCTCTATGATGCAAGACCTTGTTCATATGTTTGTCTGACCTTC 730
Db 160 ATCTCTTACATGCTCTATGATGCAAGACCTTGTTCATATGTTTGTCTGACCTTC 101
Qy 731 TCCCAACATTTGCTTGTGACCCCTGACACATCCCTCTTCGAGAAACACCCACAGATGA 790
Db 100 TCCCAACATTTGCTTGTGACCCCTGACACATCCCTCTTCGAGAAACACCCACAGATGA 41
Qy 791 TCATTAATATCTAAGGAACTCAGA 815
Db 40 TGATTAATATCTAAGGAACTCAAA 16

RESULT 6

B0575726/c 689 bp mRNA linear EST 19-JUN-2002
LOCUS UI-H-E21-bbg-f-06-0-UI.s1 NCI CGAP Ch2 Homo sapiens cDNA clone
DEFINITION UI-H-E21-bbg-f-06-0-UI 3', mRNA sequence.

ACCESSION B0575726
VERSION B0575726.1 GI:21479043
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 689)
NCI-GAP <http://www.ncbi.nlm.nih.gov/ncigap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of
Orthopedics
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 16-689, >LTR5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

Location/Qualifiers
1..689
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-E21-bbg-f-06-0-UI"
/tissue_type="Chondrosarcoma Grade II"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Ch2"
/note="Organ: Left Pelvis; Vector: pRTT3-Pac (Pharmacia)
with a modified polylinker; Site 1: Ecor I; Site 2: Not I;
NCI CGAP Ch2 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma Grade II. The library
was constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA

synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is TGATCAGCT.
TAG TISSUE=grade-2-chorioidsarcoma
TAG LIB=UI-H-E21
TAG_SEQ=ATCTAATATG"

ORIGIN

Query Match 71.6%; Score 629.8; DB 13; Length 689;
Best Local Similarity 97.6%; Pred. No. 2.9e-181;
Matches 659; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

QY 142 GTGCTCTCTGAAACATGCTGTGTCACTCAAGGTTAATGATTAAGGCGGTGACAG 201
DB 689 GTGCTCTCTGAAAGTGCTGTGTCACTCAAGGTTAATGATTAAGGCGGTGACAG 630
QY 202 ATGAGCTTTGTTAAAGATGCTTGAAGGAGCATGCTCTTAAGATCATCACCCTCC 261
DB 629 ATGAGCTTTGTTAAAGATGCTTGAAGGAGCATGCTCTTAAGATCATCACCCTCC 570
QY 262 CTAACTCAAGTACCAAGGACACAAAACCTGCGAAAGCGGAGGACCTCTGCTAGG 321
DB 569 CTAACTCAAGTACCAAGGACACAAAACCTGCGAAAGCGGAGGACCTCTGCTAGG 511
QY 322 AAAGCCAGATTTTCCAGCTTTCTCCCATGTGATAGCTGAAATATGCGCTCTGGG 381
DB 510 AAAGCCAGATTTTCCAGCTTTCTCCCATGTGATAGCTGAAATATGCGCTCTGGG 451
QY 382 AAAGGAAGACCTGACCGTCCCGAGCCGACACCCGTAAGAGGCTGTGTGAGAGGA 441
DB 450 AAAGGAAGACCTGACCGTCCCGAGCCGACACCCGTAAGAGGCTGTGTGAGAGGA 391
QY 442 TTAAGTAAAGAGAGAGATGCTCTTGTGAGATGAGCAAGAGAGAGCATCTGTCTCT 501
DB 390 TTAAGTAAAGAGAGAGATGCTCTTGTGAGATGAGCAAGAGAGAGCATCTGTCTCT 331
QY 502 GCGTGTCCCTGGGAGATGATGCTGGTATAAACCCTGTTGATGCTCATCTACTG 561
DB 330 GCGTGTCCCTGGGAGATGATGCTGGTATAAACCCTGTTGATGCTCATCTACTG 271
QY 562 AGATAGGAGAAACCGGCTTGAAGGCTGAGAGTGGAGCTGCGGAGAGCAATGCTTTG 621
DB 270 AGATAGGAGAAACCGGCTTGAAGGCTGAGAGTGGAGCTGCGGAGAGCAATGCTTTG 211
QY 622 TAAAGCATGATGTTATGTTATGATCATCTTAAAGCAGACACTTAATCTTTACA 681
DB 210 TAAAGCATGATGTTATGTTATGATCATCTTAAAGCAGACACTTAATCTTTACA 151
QY 682 TTGTCTATGATGAAAGACCTTTGTTCATGATGTTGTCTGTGACCTCTCCAC-AAT 740
DB 150 TTGTCTATGATGAAAGACCTTTGTTCATGATGTTGTCTGTGACCTCTCCACAAAT 91
QY 741 TGCTTTGAGCCCTGACACATCCCTCTTGAAGAAACCCACAGATGATCAATAATA 800
DB 90 TGCTTTGAGCCCTGACACATCCCTCTTGAAGAAACCCACAGATGATCAATAATA 31
QY 801 CTAAGGAGACTCAGA 815
DB 30 CTAAGGAGACTCAGA 16

RESULT 7
CD588661
LOCUS
DEFINITION
ACCESSION
VERSION

CD588661 874 bp mRNA linear EST 11-JUN-2003
AGENCOURT 14365300 NIH MGC 181 Homo sapiens cDNA clone
IMAGE:30355730 5', mRNA sequence.
CD588661
CD588661.1 GI:315847229

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

EST.
Homo sapiens (human)
Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 874)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-rc@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDMA76 row: e column: 19
High quality sequence stop: 778.
Location/Qualifiers

FEATURES

SOURCE

1. 874
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30395730"
/tissue_type="White Matter"
/dev_stage="Unknown"
/lab_host="DH10B-Ton A (T1 and T5 phage resistance)"
/clone_lib="NIH MGC 181"
/note="Vector: pCMV-Sport6.1; Site 1: NotI; Site 2: EcoRV
(destroyed); Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.42 kb. Library was constructed by
(Invitrogen). Note: this is a NIH MGC Library."

ORIGIN

Query Match 71.3%; Score 627; DB 14; Length 874;
Best Local Similarity 96.2%; Pred. No. 2.3e-180;
Matches 653; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

QY 201 GATGCTTTGTTAAACAGATGCTTGAAGGAGCAGATGCTCTTAAGATCATCACTC 260
DB 91 GATGCTTTGTTAAACAGATGCTTGAAGGAGCAGATGCTCTTAAGATCATCACTC 150
QY 261 CTTAATCTCAAGTACCAAGGAGACAAAACCTGCGAAGGCGGAGGACCTTGCTAG 320
DB 151 CTTAATCTCAAGTACCAAGGAGACAAAACCTGCGAAGGCGGAGGACCTTGCTAG 210
QY 321 GAAAGCAGTATTTGCCAGCTTTTCTCCCATGTGATACCTGAAATATGCGCTCGTG 380
DB 211 GAAAGCAGTATTTGCCAGCTTTTCTCCCATGTGATACCTGAAATATGCGCTCGTG 270
QY 381 GAAAGGAAAGCTGACCGTCCCGAGCCGACACCCGTAAGAGGCTGTGTGAGAGG 440
DB 271 GAAAGGAAAGCTGACCGTCCCGAGCCGACACCCGTAAGAGGCTGTGTGAGAGG 330
QY 441 ATTAGTAAAGAGAGAAATGCTCTTGAAGTTGAGACAAGAGAGCATCTGTCTCC 500
DB 331 ATTAGTAAAGAGAGAAATGCTCTTGAAGTTGAGACAAGAGAGCATCTGTCTCC 389
QY 501 TGCTGTCCCTGGGCAATGAAATGCTCGGTATTAACCCGATTTGATGCTCATCTACT 560
DB 390 TGCTGTCCCTGGGCAATGAAATGCTCGGTATTAACCCGATTTGATGCTCATCTACT 449
QY 561 GAGATAGGAGAAACCGCTTGAAGGCTGAGAGTGGAGCTGCGGAGCAATATCTGCTTT 620
DB 450 GAGATAGGAGAAACCGCTTGAAGGCTGAGAGTGGAGCTGCGGAGCAATATCTGCTTT 509
QY 621 GTAAGCATGAGATGTTATGTTATGATCATCTTAAAGCAGACGACTTAATCTTTAC 680

Db 510 GTAAGCATGAGATGTTATGTATGATGATATCTAAAGTAGACACTTAATCCTTAC 569
Qy 681 ATTGTCATATGATGAAAGACCTTTGTTTCAATGTTTGTCTGTACACCTCTCCCAACAT 740
Db 570 CTGTCTATATGTGAAAGACCTTGTGTACGTGTTGTCTGTACCTCTCCCAACAT 629
Qy 741 TGTCTTGTGACCCGTGACATCCCTCTTGTGAGAAACACCCACAGATGATCAATAATA 800
Db 630 TGTCTTGTGACCCGTGACATCCCTCTTGTGAGAAACACCCACAGATGATCAATAATA 689
Qy 801 CTAAAGGAACTCAGAGGCTGGCGGATCTTCATATGCTGAACGCTGGTTCCCGGATCC 860
Db 690 CTAAAGGAACTCAGAGGCTGGCGGATCTTCATATGCTGAACGCTGGTTCCCGGATCC 749
Qy 861 CTTCTTCTTCTTCTCTATA 879
Db 750 CTTATTTCTTCTCTATA 768

RESULT 8
B0216036/c 818 bp mRNA linear EST 02-MAY-2002

LOCUS B0216036 7553694 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6046113
DEFINITION 5', mRNA sequence.

ACCESSION B0216036
VERSION B0216036.1 GI:20397447

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 818)
NHL-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgabs@emil.nih.gov

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LLM13291 Row: 9 Column: 10

High quality sequence stop: 577.

Location/Qualifiers

1..818

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6046113"

/release_type="embryonal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_1lb="NIH_MGC_92"

/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NCI;
Site 2: Sall; Cloned unidirectionally; Oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN

Query Match 70.8%; Score 622.4; DB 13; Length 818;
Best Local Similarity 97.5%; Pred. No. 5.8e-179;
Matches 632; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 69 CCATTTGTTATGATGAAAGAAATCTTCTGCTTGAATTTCTGTAATGATGACT 128
Db 648 CCATTTGTTATGATGAAAGAAATCTTCTGCTTGAATTTCTGTAATGATGACT 589
Qy 129 TACCCCAACCCCGTGTCTCTGAAACATGTGTGTCCACTCAGGGTTAAATGATTA 188

Db 588 TACCCCAACCCCGTGTCTCTGAAACATGTGTGTGTCACTCAGATTAAATGATTA 529
Qy 189 AGGGCGGTGAGAGATGTGCTTGTGTTAAACAGATGCTTGAAGGAGAGATGCTTAAAG 248
Db 528 AGGGCGGTGAGAGATGTGCTTGTGTTAAACAGATGCTTGAAGGAGAGATGCTTAAAG 469
Qy 249 TCATACACATCCCTTAATCTCAATGATCCAGGAGACAAAATCGGAGAGCGCAGG 308
Db 468 TCATACACATCCCTTAATCTCAATGATCCAGGAGACAAAATCGGAGAGCGCAGG 409
Qy 309 ACTCTGCTTGAAGAGCAGATATTTGCAAGCTTTTCCCATGTGATGCTGAAT 368
Db 408 ACTCTGCTTGAAGAGCAGATATTTGCAAGCTTTTCCCATGTGATGCTGAAT 349
Qy 369 ATGGCTGTGAG 428
Db 348 ATGGCTGTGAG 289
Qy 429 GTGCTGAGAGAGATTAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
Db 288 GTGCTGAGAGAGATTAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 229
Qy 489 GCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548
Db 228 GCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 169
Qy 549 GCTTCATCTACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608
Db 168 GCTTCATCTACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 109
Qy 609 CAATCTGCTTGTGAAAGCACTGAGATGTTATGATGATGATGATGATGATGATGATG 668
Db 108 CAATCTGCTTGTGAAAGCACTGAGATGTTATGATGATGATGATGATGATGATGATG 49
Qy 669 TTAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 716
Db 48 TTAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1

RESULT 9
BUI98923 962 bp mRNA linear EST 05-SEP-2002

LOCUS BUI98923 DCB Homo sapiens cDNA, mRNA sequence.

DEFINITION BUI98923

ACCESSION BUI98923.1 GI:22717327

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 962)

AUTHORS Xu, X., Gu, J., Liu, F., Qu, J., Zhao, M., Li, Y., Huang, Q., Zhou, J.,
Song, H., Gu, Y., Yang, Y., Gao, G., Xiao, H., Li, N., Qian, B., Gao, X.,
Cheng, Z., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M.,
Lu, G., Cheng, Z., and Han, Z.

TITLE Homo sapiens cDNA DCB clones

JOURNAL Unpublished (2000)

COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn.

FEATURES

source

1..962

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/cell_type="dendritic cells"

/dev_stage="mature"

/lab_host="BM25.8"

/clone_1lb="DCB"

LOCUS	CD244241/c	991 bp	mRNA	linear	EST 22-MAY-2003
DEFINITION	AGENCOURT.14097645 NIH_MGC.180 Homo sapiens cDNA clone				
VERSION	IMAGE:30378143 5', mRNA sequence.				
ACCESSION	CD244241				
KEYWORDS	CD244241.1 GI:31004705				
SOURCE	EST.				
ORGANISM	Homo sapiens (human)				
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 991) NIH-MGC http://mgs.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. Michael Brownstein cDNA Library Preparation: Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: ND4443 row: h column: 24 High quality sequence stop: 728.				
FEATURES	Location/Qualifiers				
SOURCE	1..891 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:30378143" /lab_host="MD10B-Ron A. (T1 and T5 phage resistances)" /notes="Organ: Testis; Vector: pCMV-SPORT.1; Site_1: NotI, Site_2: EcoRV (destroyed); Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.68 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."				
ORIGIN					
Query Match	69.5%; Score 610.6; DB 14; Length 891;				
Best Local Similarity	93.8%; Pred. No. 2.5e-175;				
Matches	682; Conservative 0; Mismatches 34; Indels 11; Gaps 4;				
OY	1 TGTGGGGAAGAGAGAGATGATGTTACTGTGTCTGTGTGAAAGAGTACAT	60			
DB	717 TGTGGGGAAGAGAGAGATGATGTTACTGTGTGTGTGAAAGAGTACAT	658			
OY	61 AGGAGACTCCATTGTTATGTAATGACAAAGAAATCTTGCTGAGATTCTGTAATC	120			
DB	657 AGGAGACTCCATTGTTATGTAATGACAAAGAAATCTTGCTGAGATTCTGTAATC	598			
OY	121 TATGACCTTACCCCAACCCCGTGTCTCTGAAACATGTGCTGTCCACTCAGGGTTAA	180			
DB	597 TATAACTTACCCCAACCCCGTGTCTCTGAAAC--GTGCTGTGTCCACTCAGGGTTAA	540			
OY	181 ATGATTAAAGGCGGTGACAGATGTGCTTTGTTAAACAGATCTTGAAGGACGATGCTC	240			
DB	539 ATGATTAAAGGCGGTGACAGATGTGCTTTGTTAAACAGATCTTGAAGGACGATGCTC	480			
OY	241 CTTAGAGCATCACACCTCCCTAATCTCAAGTACCCAGGAGACAAAACTGCGGAAGG	300			
DB	479 CTTAGAGCATCACACCTCCCTAATCTCAAGTACCCAGGAGACAAAACTGCGGAAGG	420			
OY	301 CCGCAGGACCTCTGCTAGGAAGCCAGGTATTGTCCAACTTTCTCCCATGTGATG	360			
DB	419 CCTCAGGACCTCTGCTAGGAAGCCAGGTATTGTGTCCAACTTTCTCCCATGTGATG	360			
OY	361 CCTGAATATGTGGCTCTGTGGAGGAAAGGAAAGACTCAGCTCCCCAGGCCGACCCGTA	420			

Db	359	TCGTAAATATGCGCTCTCTGGGAATGAAAGACCTGACCGTCTCCCAACCGACGCCATA	300
Qy	421	AAGGCTGTGCTGAGGAGGATTAGTAAAGAGAGAAATGCCTC-TTGCACTTGAGAC	479
Db	299	AAGGCTGTGCTGAGGAGCATTAGTATTAAGAGAGAAATGCCTCTTTGCACTTGAGAC	240
Qy	480	AAGAGGAAGCATTCTGTCTCTGCTGCTCCCTGGGCAATGGAATGTCGGGTATAAACC	539
Db	239	AAGGGAATCATCTGTCTCTGCTGCC-GTGCCTGGCAATGGAATGTCTCGATATAAACC	181
Qy	540	CGATTGTATGCTCCATCTTACTGAGATAGGAAAAACCGCTTAGGCGTGAAGTGGAGCC	599
Db	180	CGATTGTATGCTCATCTTACTGAGATAGGAAAAACCGCTTAGGCGTGAAGTGGAGAAC	121
Qy	600	TGCGGAGCAATCTGCTTTGTAAAGCACTGAGATGTTATGTATGCAATATCTAAA	659
Db	120	T-----GCAATCTCTTTGTAAACCAATGAGATGTTATGTATGCAATATCTAAA	68
Qy	660	GCACAGCACTTAATCCCTTTACATGTCTATGAGCAAGACCTTGTTGACATGTTGTC	719
Db	67	GCACAGCACTTAATCTTTACCTTTGATGATGATGCAAGACCTTGTGTACGTTGTGC	8
Qy	720	TGCTGAC	726
Db	7	TGCTGAC	1
RESULT 12			
CB850840/c		670 bp	mRNA
LOCUS		UI-CF-EN1-ada-a-13-0-UI.91	UI-CF-EN1 Homo sapiens cDNA clone
DEFINITION		UI-CF-EN1-ada-a-13-0-UI 3', mRNA sequence.	
ACCESSION		CB850840	
VERSION		CB850840.1	GI:30045609
KEYWORDS		EST.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE		1 (bases 1 to 670)	
JOURNAL		Ronald M.F., Lennon G. and Soares M.B.	
MEDLINE		Normalization and subtraction: two approaches to facilitate gene	
PUBMED		discovery	
COMMENT		Genome Res. 6 (9), 791-806 (1996)	
		97044477	
		8889548	
		Contact: McCray, PB	
		McCray Lab	
		University of Iowa	
		2024 University of Iowa Med Labs, Iowa City, IA 52242, USA	
		Tel: 319 356 4866	
		Fax: 319 356 7171	
		Email: paul-mccray@iowa.edu	
		Tissue Procurement: Dr. M. J. Welsh, University of Iowa	
		cDNA library preparation: Dr. M. Bento Soares, University of Iowa	
		DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa	
		cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa	
		Clone Distribution: Researchers may obtain clones from Research	
		Genetics (www.resgen.com).	
		The following repetitive elements were found in this cDNA	
		sequence: 18-670, S1R5#LTR/Retroviral (matched complement)	
		Seq primer: M13 FORWARD	
		POLYA=yes.	
FEATURES			
source			
		Location/Qualifiers	
		1..670	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/cclone="UI-CF-EN1-ada-a-13-0-UI"	
		/tissue_type="Primary Lung Cystic Fibrosis Epithelial	
		cells"	
		/dev_stage="Adult"	

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.
TAG_SEQ=None found"

ORIGIN

Query Match 68.1%; Score 598.8; DB 14; Length 670;
Best Local Similarity 95.7%; Pred. No. 9e-172;
Matches 626; Conservative 0; Mismatches 27; Indels 1; Gaps 1;

QY 226 GAAGGACAGATGCTCTTAAGAGTCATCCACTCCCTAATTCAGTACCAGGACAC 285
DB 670 GAAGGACAGATGCTCTTAAGAGTCATCCACTCCCTAATTCAGTACCAGGACAC 611
QY 286 AAAAAGCGGAGGAGCGGAGGAGCTCTGCTGAGGAAAGCAAGGATTCACAGTTT 345
DB 610 AAAACATCGCGAAGGCGGAGGAGCTCTGCTGAGGAAAGCAAGGATTCACAGTTT 551
QY 346 CTCCTCCATGATAGCTCTGAATATGAGCTCTGAGGAAAGGAAAGACCTGATCC 405
DB 550 CTCCTCCATGATAGCTCTGAATATGAGCTCTGAGGAAAGGAAAGACCTGATCC 491
QY 406 AGCCCGACACCCGTAAGAGGTCTGTGCTGAGAGGATTAATAAGAGAAAGATCTT 465
DB 490 TGCCTCGACACCCGTAAGAGGTCTGTGCTGAGAGGATTAATAAGAGAGATCTT 431
QY 466 CTGCGAGTTAGACAAAGAGGAGCATGCTCTGCTGAGGAAAGGAAAGGAAATGT 525
DB 430 CTGCGAGTTAGACAAAGAGGAGCATGCTCTGCTGAGGAAAGGAAAGGAAATGT 371
QY 526 CTGCGTATAAAGCCGATGTTATGCTCCATCTAGATAGGAAAGAAAGCCGCTTAG 585
DB 370 CTGCGTATAAAGCCGATGTTATGCTCCATCTAGATAGGAAAGAAAGCCGCTTAG 312
QY 586 CTGAGAGTGGAGCTGCGGAGCAATATGCTTTGTAAGCACTGAGATGTTATGT 645
DB 311 CTGAGAGTGGAGCAATGCGGAGCAATATGCTTTGTAAGCAATGATGTTATGT 252
QY 646 ATGATATCTAAAGACAGCACTTAATCTTAATGATGATGATGCAAGGACCTTG 705
DB 251 ATGATATCTAAAGACAGCACTTAATCTTAATGATGATGATGCAAGGACCTTG 192
QY 706 TTCAATGTTTGTCTGTCGACCTCTCCCAATTTCTTGTGAGACCTGACATCCC 765
DB 191 TTCAATGTTTGTCTGTCGACCTCTCCCAATTTCTTGTGAGACCTGACATCCC 132
QY 766 CTCTTGAGAAACCAACAGATGATCAATTAATACTAAGGAACTGAGAGCTGGCG 825
DB 131 CTCTTGAGAAACCAACAGATGATCAATTAATACTAAGGAACTGAGAGCTGGCG 72
QY 826 ATCTCTCATATGTCGAGAGCTGCTGCTCCCGGCTCCCTTTCTTCTCTATA 879
DB 71 ATCTCTCATATGTCGAGAGCTGCTGCTCCCGGCTCCCTTTCTTCTCTATA 18

RESULT 13
B0006641/c 670 bp mRNA linear EST 26-MAR-2002
LOCUS UI-H-E11-ayz-b-08-0-UI.s1 NCI_CGAP_E11 Homo sapiens cDNA clone
DEFINITION IMAGE:5845543 3', mRNA sequence.

ACCESSION B0006641 GI:19731541
VERSION B0006641.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 670)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
The following repetitive elements were found in this cDNA
sequence: 18-670 >LTR5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=tes.

FEATURES

Location/Qualifiers
source . 1..670
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5845543"
/issue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="NCI CGAP E11"
/clone_lib="NCI CGAP E11"
/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP_E11 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTTGCAC.
TAG_TISSUE=chondrosarcoma
TAG_LIB=UI-H-E11
TAG_SEQ=AACTTGCAC"

ORIGIN

Query Match 68.0%; Score 597.6; DB 12; Length 670;
Best Local Similarity 94.6%; Pred. No. 2.1e-171;
Matches 618; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 227 AAGGACAGATGCTCTTAAGAGTCATCCACTCCCTAATTCAGTACCAGGACACA 286
DB 670 AAGGACAGATGCTCTTAAGAGTCATCCACTCCCTAATTCAGTACCAGGACACA 611
QY 287 AAAAAGCGGAGGAGCGGAGGAGCTCTGCTGAGGAAAGCAAGGATTCACAGTTT 346
DB 610 AAAAAGCGGAGGAGCGGAGGAGCTCTGCTGAGGAAAGCAAGGATTCACAGTTT 551
QY 347 TCCCATGTGATAGCTGAATATGAGCTCTGAGGAAAGGAAAGCACTGACCTCC 406
DB 550 TCCCATGTGATAGCTGAATATGAGCTCTGAGGAAAGGAAAGCACTGACCTCC 491
QY 407 GCCCGACACCCGTAAGAGGTCTGTGCTGAGAGGATTAATAAGAGGAAAGATGCTC 466
DB 490 GCCCGACACCCGTAAGAGGTCTGTGCTGAGAGGATTAATAAGAGGAAAGATGCTC 431

OY	467	TTGCAGTTGAGACAAGAGAAGGACATCTGTCTCCCTGCCTGTCGTCCTCGGGCAATYGAATATTC	526
Db	430	TTGCAGTTGAGACAGAGAGAGAGGACATCTGTCTCTCGCCCGTCCCTGGGCAATGGAATATTC	371
OY	527	TCGGTATATAAACCCCGATTGTATGCTCATCTACTAGATATAGGAAAAACCGCTTAAAGGC	586
Db	370	TCGGTATATAAACCCCGATTGTACGTTCCATCTACTAGATATAGGAAGAAAAACGCTTAAAGGC	311
OY	587	TGAGAGTGGGACCTGCGGGCAGACCAATACTGCTTTGTGTAAGACACTGAGATGTTTATGTTGA	646
Db	310	TGAGAGTGGGACATGCGAGCAGACCAATACTGCTTTGTGTAAGACACTGAGATGTTTATGTTGA	251
OY	647	TGCATATCTAAAAAGACAGACACTTATCCTTACATGTGCTATGATGAGAAAGACCTTTGT	706
Db	250	TGCATATCTAAAAAGACAGACACTTATCCTTACATGTGCTATGATGAGAAAGACCTTTGT	191
OY	707	TCACATGTTTTGTCTGCTGACCCCTCTCCGCAACAATGTTCTGTGACCTTGACATATCCCC	766
Db	190	TCACATGTTTTGTCTGCTGACCCCTCTCCGCAACAATGTTCTGTGACCATGACATATCCCC	131
OY	767	TCTTGAAGAAACCCACAGATGATCAGTAAATACTAAAGGAACTCAAGAGCTGGCGGGA	826
Db	130	TCTCAGAGAAACACCCACAGATGATCAATAAATACTAAAGGAACTCAAGAGCGGGCGGA	71
OY	827	TCCTTCATATGCTGAAGCTGGTTGCCCGGGGATCCCTTCTTTCTTCTCTATA	879
Db	70	TCCTTCATATGCTGAAGCTGGTTGCCCGGGGATCCCTTCTTTCTTCTCTATA	18

RESULT	14
BM68693/c	
LOCUS	BM68693
DEFINITION	633 bp mRNA linear EST 27-FEB-2002 UI-E-CX1-afm-b-10-0-UI .g2 UI-E-CX1 Homo sapiens cDNA clone UI-E-CX1-afm-b-10-0-UI 3', mRNA sequence.

ACCESSION	BM668693
VERSION	BM668693.1
KEYWORDS	GI:18978590
SOURCE	EST.
	Homo sapiens (human)

ORGANISM	Home sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
REFERENCE	
1 (bases 1 to 633)	

AUTHORS Ronaldo, M. F., Lennon, G. and Soares, M. B.
TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477
 PUBMED 8889548
 COMMENT
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics

University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com).
The following repetitive elements were found in this cDNA sequence: 18-633...LTR5'LTR/Retroviral (matched complement)
Seq primer: M13 Forward
POLYA=yes.

FEATURES	Location/Qualifiers
source	1. .633

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="caxon:9606"
/clonetype="UI-E-CR1-afm-b-10-0-UI"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"

```

```

/abc host="DH10B (Life Technologies) (T1 phage resistant)
/clone lib="ui-E-CX1"
./note:Organ: eye; Vector: pT73-pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
ui-E-CX1 is a normalized cDNA library containing the
following tissue(s): Retina Foveal and Macular. The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an Ecor I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GTCC. This library was created for the program,
Gene Discovery in the Visual System, supported by National
Eye Institute (NEI).
TAG_TISSUE=Foveal and Macular Retina
TAG_LIB=ui-E-CX1
TAG_SEQ=GTCC"

```

Query Match	67.9%	Score 597;	DB 12;	Length 633;
Best Local Similarity	97.6%;	Pred. No. 3,1e-171;		
Matches 603; Conservative	0;	Mismatches 15;	Indels 0;	Gaps 0;

262 CTAATCTCAAGTACCCAGGACACAAAACTGCGAAGGCCGACGGACCTTCGCTTAGG 321

Db 633 CTAATCTCAAGTACNCAAGNACCAAAACTGCCGAGGCCCGACAGACCTCTGCNTAGG 574

322 AAAGCAGGATGTTCGAACGTTCTCTCCCATGTGATAGCCTGAAAATATGCCCTCGTGGG 381

Db 573 AAAGCAGGTAATTCTCCAGGTTCTCCCCATGATAGTCTGAATATGCGCTCGTGG 514

382 AAGGAAGACCTGACCGTCCCCAGCCCGACACCCGTAAGGCTGTGTGAGAGAGA 441

Db 513 AAGNGAAGACTGACCGTCCCCCAGCCGACACCCGTAAGGGTCTGTGCTGAGGAGGA 454

442 TTAGTAAAGAGAGAGGATGCTCTTTCAGTTGAGACAAAGAGAAAGCATCTGTCTCT 501

Db 453 TTAGTAAAGAGGAAATGCTCTTGCAAGTTGAGACAAAGAGGAAAGGCATCTGTCTTCT 394

502 GCCTGTCCTCCGCGCAATGGAAATGTCCTGGTATATAAAACCCGATTGTAATGCTCCATCTACTG 561

Db 393 GCCTGTCCTGGGCAATGGAATGTCCTGGATATAAACCCGATTGTATGCTCCATCTACTG 334

562 AGATAGGGAAAAACCGCCTTAGAGGCTGAGGTGGGACCTGCGGGCAGCAATATGCTTTG 621

Db 333 AGATAGGGAAAAACCGCCTTAGGGCTGGAGTGGGACCTCGGGGGCAGCAATACGCTTTC 214

622 TAAAGCACTGAGATGTTATGTCATGATCTAAAGCA CAGCACTTAATCCCTTACA 681

DB 2/3 1AAAGCAGCAGAGATGGTATAGTATGCAATATCCAAAAGCAGCAGC11AAATCCCTTACCA 214

08Z 11G1C1A1GAI9CAAGAC111G11CA6A1G111G11C15F1GAACCC1C1CCCCCAC6611742

213 11G1C1A1GAI9CAAGAC111G11CA6A1G111G11C15F1GAACCC1C1CCCCCAC6611742

213 1181C1A15915C8F86AC111811C86E911181C10C158F8CC1C10C8C8A11 12

153 GTCTTGGACCCCTGACACATCCCCCTCTTTGGAGAAACACCCACGGATGATCATTAATAC 94

802 TAAAGGA^{CT}CAGAGGCTGGCGGGATCCTCCATAATGCTGAACGCTGGTTC^{CT}CCCGGGTCCC 861

Db 93 TAAGGAACTCAGAGGCTGGCGGATCCCTTCATATGCTGAACGCTGTTCCCCGGGTCCC 34

862 CTCCTTCTTCTCTATA 879

Db 33 CTATTCTTCTCTAA 16

RESULT 15

CA449778/c 678 bp mRNA linear EST 08-NOV-2002
LOCUS
DEFINITION
UI-H-E11-azd-1-19-0-UI.s1 NCI CGAP E11 Homo sapiens cDNA clone
UI-H-E11-azd-1-19-0-UI 3', mRNA sequence.
ACCESSION
CA449778
VERSION
CA449778.1 GI:24814198
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 678)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
COMMENT
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 16-678, >L1R5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

source

Location/Qualifiers
1..678
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-E11-azd-1-19-0-UI"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_id="NCI CGAP E11"
/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP E11 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma. The library was
constructed according to Bonaldo, Lennom and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
ACACTTGCAC.
TAG_TISSUE=chondrosarcoma
TAG_LIB=UI-H-E11
TAG_SEQ=ACACTTGCAC"

ORIGIN

Query Match 67.5%; Score 593.4; DB 14; Length 678;
Best Local Similarity 95.1%; Pred. No. 4,1e-170;
Matches 634; Conservative 0; Mismatches 31; Indels 2; Gaps 2;
QY 213 TAAACAGATGCTTGAAGCAGATGCTCTTAAGATCATCAGCATCCCTAATCTCAAG 272
DB 678 TAAACAGATGCTTGAAGCAGATGCTCTTAAGATCATCAGCATCCCTAATCTCAAG 619
QY 273 TACCCAGGACACAAATCTCGGAGAGCCGACAGGACCTTCTGCTTAGGAAAGCAGGTA 332
DB 618 TACCCAGGACACAAATCTCGGAGAGCCGACAGGACCTTCTGCTTAGGAAAGCAGGTA 559
QY 333 TTGTCCAAAGCTTTCTCCCATGTGATAGCCTGAATATAGGCTCTGTGGAAAGGAAAGAC 392
DB 558 TTGTCCAAAGCTTTCTCCCATGTGATAGCTGAATATAGGCTCTGTGGAAAGGAAAGAC 499

QY 393 CTGACCGTCCCCAGAGCCGACACCCGTTAAAGGCTCTGTGTGAGAGGATTAATAAAGA 452
DB 498 CTGACCGTCCCCAGAGCCGACACCCGTTAAAGGCTCTGTGTGAGAGGATTAATAAAGA 439
QY 453 GGAAGGAATGCTCTTGTGAGAGGATTAAGAGAGGAGGATCTCTCTCTGCTGCTG 512
DB 438 GGAAGGAATGCTCTTGTGAGAGGATTAAGAGAGGAGGATCTCTCTCTGCTGCTG 379
QY 513 GGAATGGAATGCTCTGAGTAAAGCCGATTTGATGCTCATCTAAGATAGGAAA 572
DB 378 GGAATGGAATGCTCTGAGTAAAGCCGATTTGATGCTCATCTAAGATAGGAAA 319
QY 573 AACCGCTTATGAGGCTGAGGTGGACCTGCGGACCAATACGCTTTGAAAGCACTGA 632
DB 318 AACCGCTTATGAGGCTGAGGTGGACCTGCGGACCAATACGCTTTGAAAGCACTGA 259
QY 633 GATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 692
DB 258 GATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 200
QY 693 GCAAGACCTTTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 752
DB 199 GCAAGACCTTTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 140
QY 753 CTGACATGCTCTCTGAGAGAAACCCGACGATGATGATGATGATGATGATGATGAT 812
DB 139 GTGACATGCTCTCTGAGAGAAACCCGACGATGATGATGATGATGATGATGATGAT 80
QY 813 AGAGGCTGCGGAGATCCTCATATGCTGAACGCTGCTGCCCGGATCCCTCTTCTT 872
DB 79 AGAGGCTGCGGAGATCCTCATATGCTGAACGCTGCTGCCCGGATCCCTCTTCTT 21
QY 873 CTCTATA 879
DB 20 CTATAA 14

Search completed: February 28, 2004, 01:07:13
Job time: 3579.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 22:08:49 ; Search time 625.802 Seconds

(without alignments)
10389.001 Million cell updates/sec

Title: US-10-016-604-155

Sequence: 1 gagatagggagaaacgcgcct.....atgcaagaccttcttcac 150

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	150	100.0	150	6 AX594077	AX594077 Sequence
2	150	100.0	319	6 AX593925	AX593925 Sequence
3	150	100.0	408	6 AX593926	AX593926 Sequence
4	150	100.0	879	6 AX593927	AX593927 Sequence
5	150	100.0	968	6 AX594072	AX594072 Sequence
6	150	100.0	1505	14 HERVCORF	X82271 Human endog
7	150	100.0	3366	14 HERVCORF	X82272 Human endog
8	150	100.0	8300	14 HEN17834	Y17834 Human endog
9	150	100.0	8708	14 AF490464	AF490464 Homo sapi
10	150	100.0	9463	14 AY037929	AY037929 Homo sapi
11	150	100.0	9472	14 AF164614	AF164614 Homo sapi
12	150	100.0	10029	14 HEN17832	Y17832 Human endog
13	150	100.0	19221	9 AF074086	AF074086 Homo sapi
14	150	100.0	154037	9 AC072054	AC072054 Homo sapi
15	150	100.0	175202	9 AC134684	AC134684 Homo sapi
16	150	100.0	181916	2 AC110927	AC110927 Homo sapi
17	150	100.0	188884	9 AC130367	AC130367 Homo sapi
18	150	100.0	205317	9 AC093856	AC093856 Homo sapi
19	148.4	98.9	521	9 AF276875	AF276875 Homo sapi
20	148.4	98.9	769	9 AF017229	AF017229 Homo sapi
21	148.4	98.9	996	9 AF394944	AF394944 Homo sapi
22	148.4	98.9	1156	9 AK075337	AK075337 Homo sapi
23	148.4	98.9	1877	9 AK098742	AK098742 Homo sapi
24	148.4	98.9	2431	6 BD127902	BD127902 Primer fo
25	148.4	98.9	2431	9 AK075315	AK075315 Homo sapi
26	148.4	98.9	2517	9 AK090528	AK090528 Homo sapi
27	148.4	98.9	6206	9 AF261945	AF261945 Homo sapi
28	148.4	98.9	8807	14 HEN17833	Y17833 Human endog
29	148.4	98.9	9179	9 AF164609	AF164609 Homo sapi
30	148.4	98.9	9181	9 AF164611	AF164611 Homo sapi
31	148.4	98.9	9422	9 AF164615	AF164615 Homo sapi
32	148.4	98.9	9445	9 AF164612	AF164612 Homo sapi
33	148.4	98.9	20540	9 AC010508	AC010508 Homo sapi
34	148.4	98.9	35414	6 BD243619	BD243619 TNFR rela
35	148.4	98.9	35414	9 U73641	U73641 Human Chrom
36	148.4	98.9	39935	9 AC093231	AC093231 Homo sapi
37	148.4	98.9	41477	9 AC008996	AC008996 Homo sapi
38	148.4	98.9	45092	2 AC130487	AC130487 Homo sapi
39	148.4	98.9	45995	9 AC112702	AC112702 Homo sapi
40	148.4	98.9	59688	2 AC130284	AC130284 Homo sapi
41	148.4	98.9	73221	2 AC022584	AC022584 Homo sapi
42	148.4	98.9	82367	9 AL139421	AL139421 Human DNA
43	148.4	98.9	91172	9 AC008553	AC008553 Homo sapi
44	148.4	98.9	92171	9 AC004924	AC004924 Homo sapi
45	148.4	98.9	93684	9 AC109513	AC109513 Homo sapi

ALIGNMENTS

```
RESULT 1
LOCUS AX594077 150 bp DNA linear PAT 13-FEB-2003
DEFINITION Sequence 155 from Patent WO0246477.
ACCESSION AX594077
VERSION AX594077.1 GI:28375310
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Garcia,P., Hardy,S.F., Williams,L.T. and Escobedo,J.
Endogenous retroviruses up-regulated in prostate cancer
Patent: WO 0246477-A 155 13-JUN-2002;
JOURNAL
```

CHIRON CORPORATION (US)
FEATURES
Location/Qualifiers
source 1..150
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 150; DB 6; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.6e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGAGTGGACCTGGCGGACGACAAATACCTGCTT 60
DB 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGAGTGGACCTGGCGGACGACAAATACCTGCTT 60
QY 61 GTAAGACACTGAGATGTTTATGTGTATGCATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 61 GTAAGACACTGAGATGTTTATGTGTATGCATATCTAAAGACAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
DB 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
RESULT 2
AX593925 319 bp DNA linear PAT 13-FEB-2003
LOCUS AX593925
DEFINITION Sequence 3 from Patent WO0246477.
ACCESSION AX593925
VERSION AX593925.1 GI:28375184
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.
TITLE Endogenous retroviruses up-regulated in prostate cancer
JOURNAL Patent: WO 0246477-A 3 13-JUN-2002;
CHIRON CORPORATION (US)
FEATURES
source Location/Qualifiers
1..319
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 150; DB 6; Length 319;
Best Local Similarity 100.0%; Pred. No. 1.6e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGAGTGGACCTGGCGGACGACAAATACCTGCTT 60
DB 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGAGTGGACCTGGCGGACGACAAATACCTGCTT 60
QY 61 GTAAGACACTGAGATGTTTATGTGTATGCATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 61 GTAAGACACTGAGATGTTTATGTGTATGCATATCTAAAGACAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
DB 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
RESULT 3
AX593926 408 bp DNA linear PAT 13-FEB-2003
LOCUS AX593926
DEFINITION Sequence 4 from Patent WO0246477.
ACCESSION AX593926
VERSION AX593926.1 GI:28375185
KEYWORDS
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.
TITLE Endogenous retroviruses up-regulated in prostate cancer
JOURNAL Patent: WO 0246477-A 4 13-JUN-2002;
CHIRON CORPORATION (US)
FEATURES
source Location/Qualifiers
1..408
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 150; DB 6; Length 408;
Best Local Similarity 100.0%; Pred. No. 1.5e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGAGTGGACCTGGCGGACGACAAATACCTGCTT 60
DB 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGAGTGGACCTGGCGGACGACAAATACCTGCTT 60
QY 61 GTAAGACACTGAGATGTTTATGTGTATGCATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 61 GTAAGACACTGAGATGTTTATGTGTATGCATATCTAAAGACAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
DB 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
RESULT 4
AX593927 879 bp DNA linear PAT 13-FEB-2003
LOCUS AX593927
DEFINITION Sequence 5 from Patent WO0246477.
ACCESSION AX593927
VERSION AX593927.1 GI:28375186
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.
TITLE Endogenous retroviruses up-regulated in prostate cancer
JOURNAL Patent: WO 0246477-A 5 13-JUN-2002;
CHIRON CORPORATION (US)
FEATURES
source Location/Qualifiers
1..879
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 150; DB 6; Length 879;
Best Local Similarity 100.0%; Pred. No. 1.5e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCGCCTTAGGGCTGAGAGTGGACCTGGCGGACGACAAATACCTGCTT 60
DB 561 GAGATAGGAAAAACCGCCTTAGGGCTGAGAGTGGACCTGGCGGACGACAAATACCTGCTT 620
QY 61 GTAAGACACTGAGATGTTTATGTGTATGCATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 621 GTAAGACACTGAGATGTTTATGTGTATGCATATCTAAAGACAGCACTTAATCCTTTAC 680
QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
DB 681 ATTGCTATGATGCAAGACCTTTGTTTCAC 710
RESULT 5

AX594072
LOCUS AX594072 968 bp DNA linear PAT 13-FEB-2003
DEFINITION Sequence 150 from Patent WO0246477.
ACCESSION AX594072
VERSION AX594072.1 GI:28375305
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Garcia, P., Hardy, S.F., Williams, L.T. and Escobedo, J.
TITLE Endogenous retroviruses up-regulated in prostate cancer
JOURNAL Patent: WO 0246477-A 150 13-JUN-2002;
CHIRON CORPORATION (US)
FEATURES
source
1. .968
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 150; DB 6; Length 968;
Best Local Similarity 100.0%; Pred. No. 1.5e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAGTATAGGAAAAACCCCTTAGGGCTGAGAGTGGACCTGGGCGACCAATACCTCTT 60
Db 561 GAGTATAGGAAAAACCCCTTAGGGCTGAGAGTGGACCTGGGCGACCAATACCTCTT 620
Qy 61 GTAAGACACTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 621 GTAAGACACTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 680
Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 681 ATTGCTATGATGCAAGACCTTTGTTTAC 710
RESULT 6
HERVOCORF 1505 bp mRNA linear VRL 06-APR-1995
DEFINITION Human endogenous retrovirus mRNA for central open reading frame.
ACCESSION X82271
VERSION X82271.1 GI:757869
KEYWORDS open reading frame.
SOURCE Human endogenous retrovirus
ORGANISM Human endogenous retrovirus
VIRUSES; Retrovirdae.
REFERENCE
1 Lower, R., Tonjes, R.R., Korbmayer, C., Kurth, R. and Lower, J.
TITLE Identification of a Rev-related protein by analysis of spliced
JOURNAL transcripts of the human endogenous retroviruses HTDV/HERV-K
MEDLINE J. Virol. 69 (1), 141-149 (1995)
PUBMED 7983704
REFERENCE 2 (bases 1 to 1505)
AUTHORS Toenjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1994) R.R. Toenjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O.Box 1740, D-63207 Langen, FRG
FEATURES
source
1. .1505
/organism="Human endogenous retrovirus"
/mol_type="mRNA"
/specific_host="Homo sapiens, patient GH, teratocarcinoma"
/db_xref="taxon:11827"
/note="HERV-K"
gene
1. .1505
/gene="unknown orf"
misc_feature
255. .256
/gene="unknown orf"
/note="splice site"

CDS
273. .590
/gene="unknown orf"
/codon_start=1
/protein_id="CAA57722.1"
/db_xref="GI:757870"
/db_xref="SPTREMBL:Q69383"
/translation="MNPSEMRKAPPPRRRRNRNAPLTHRNKQVTSSEQMKLPSTK
AEPPTMAQLKLTQLATKYLENTKVTQTPESMILALMIVSVSAGVNSSEETATIE
NGP"
misc_feature
436
/gene="unknown orf"
/note="additional nucleotide"
512
/gene="unknown orf"
/note="additional nucleotide"
533. .534
/gene="unknown orf"
/note="splice site"
ORIGIN
Query Match 100.0%; Score 150; DB 14; Length 1505;
Best Local Similarity 100.0%; Pred. No. 1.5e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAGTATAGGAAAAACCCCTTAGGGCTGAGAGTGGACCTGGGCGACCAATACCTCTT 60
Db 1188 GAGTATAGGAAAAACCCCTTAGGGCTGAGAGTGGACCTGGGCGACCAATACCTCTT 1247
Qy 61 GTAAGACACTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 1248 GTAAGACACTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 1307
Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 1308 ATTGCTATGATGCAAGACCTTTGTTTAC 1337
RESULT 7
HERVENV 3366 bp mRNA linear VRL 30-MAR-1995
DEFINITION Human endogenous retrovirus env mRNA.
ACCESSION X82272
VERSION X82272.1 GI:757871
KEYWORDS env gene.
SOURCE Human endogenous retrovirus
ORGANISM Human endogenous retrovirus
VIRUSES; Retrovirdae.
REFERENCE
1 Lower, R., Tonjes, R.R., Korbmayer, C., Kurth, R. and Lower, J.
TITLE Identification of a Rev-related protein by analysis of spliced
JOURNAL transcripts of the human endogenous retroviruses HTDV/HERV-K
MEDLINE J. Virol. 69 (1), 141-149 (1995)
PUBMED 7983704
REFERENCE 2 (bases 1 to 3366)
AUTHORS Toenjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1994) R.R. Toenjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O.Box 1740, D-63207 Langen, FRG
FEATURES
source
1. .3366
/organism="Human endogenous retrovirus"
/mol_type="mRNA"
/specific_host="Homo sapiens, patient GH, teratocarcinoma"
/db_xref="taxon:11827"
/note="HERV-K"
gene
1. .3366
/gene="env"
CDS
437. .2536
/gene="env"
/codon_start=1
/protein_id="CAA57723.1"
/db_xref="GI:757872"

```

/db_xref="GOA:O69384"
/db_xref="SPTREMBL:O69384"
/translation="WNPSEWQRKAPRRRRRRNRAPLTHTKNTKNTVISEBQKLPSTYKK
AEPRTMAQKKLTQLATKYLENTYVTVTPRESMLLAALMIVSMVSLPMPAGAAANTYT
YMAVYPPPLIRAVTWMNDPTEVYVNVSVWPGPIDRCPAKPEEGEMINI SIGYHY
PKICGARGCGELMAPQWMLVEVPVSPICAPTYHAMSGSLRRNVYMLQDFSYORSL
KFRKGRCPKEIKPRESKENTVLWEECVANSAVILNNNEGGTIDMAPGQFPHNSL
GQTOCSBQAQVSPAVDSLTLSLKHKKKQLQSYYPMEWKGISTREPKTVSPVSPGR
EHPMLRLTVASHHRTIRISGQITLRTKRPFTITLNSSLTVLQSCVKEPYMLVVG
NIVLKPSPQITTCENCRLLCTIDSTFPMQHRILIVRAEGVMIPIVSDRPEASPVH
ILTVLKGVLNRSKRFLFTLLAVIMGLAATAVAAVAGVALHSVGSVNFVMDQKNS
TRLNNGSSIDOKLANQINDPRTQVIMWGDMLSEHRFQLOCMNTSDFCITQIYNV
ESEHHMNVRRHLQGRENDLTLDISKLKEQIFEASKAHILNVPCTEALAGAGDGLANL
NPYTWVKTIGSTTINLLILVLVCLFLCLLVRCYQQLRRDSHREARMYAVLSKKR
GNGNGSKSRDQIVTVSV"
misc_feature
600
/gene="env"
/note="additional nucleotide"
676
/gene="env"
/note="additional nucleotide"

ORIGIN
Query Match 100.0%; Score 150; DB 14; Length 3366;
Beet Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GAGATAGGGAACAAACCCGCTTAGGAGCTGGAAGTGGACCTGCGGAGCAATACCTGCTT 60
Db 3051 GAGATAGGGAACAAACCCGCTTAGGAGCTGGAAGTGGACCTGCGGAGCAATACCTGCTT 3110

Oy 61 GTAAAGCCTGAGATGTTTATGTGTATGCATATCTTAAAGACAGACCTTAATCCTTTAC 120
Db 3111 GTAAAGCCTGAGATGTTTATGTGTATGCATATCTTAAAGACAGACCTTAATCCTTTAC 3170

Oy 121 ATGTCTATGATGCAAGACCTTGTTCAC 150
Db 3171 ATGTCTATGATGCAAGACCTTGTTCAC 3200

RESULT 8
HEN17834 8300 bp DNA linear VRL 21-OCT-1999
DEFINITION Human endogenous retrovirus K (HERV-K) elements, clone SD-C7-34LTR.
ACCESSION Y17834
VERSION Y17834.1 GI:4185945
KEYWORDS env gene; env protein; gag gene; gag protein; long terminal repeat;
pol gene; pol protein.
SOURCE Human endogenous retrovirus K
ORANISM Human endogenous retrovirus K
VIRUSES; Retrovirdae.
REFERENCE
1
Tonjes,R.R., Czanderma,F. and Kutth,R.
Genome-wide screening, cloning, chromosomal assignment, and
expression of full-length human endogenous retrovirus type K
J. Virol. 73 (11), 9187-9195 (1999)
JOURNAL MEDLINE 99445825
PUBMED 10516026
2 (bases 1 to 8300)
Tonjes,R.R.
Direct Submission
Submitted (22-JUL-1998) R.R. Tonjes, Paul-Ehrlich Inst.,
Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-63207 Langen, FRG
COMMENT Related sequence Y10390.
FEATURES
Location/Qualifiers
1..8300
/oranism="Human endogenous retrovirus K"
/proviral
/mol_type="genomic DNA"
/specific_host="human"
/db_xref="taxon:45617"
/clone="SD-C7-34LTR"
/note="assigned to human chromosome 7"

```

gene
CDS
'
32. . 2032
/gene="gag"
32. . 2032
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="CA76884.1"
/db_xref="GI:1185946"
/db_xref="GOA:O9YNA6"
/db_xref="SPTRMBL:O9YNA6"
/translation="MGQTGSKIKSVKASVSYFIKILKRGVKNVSTNLIKLFQILEQ
FCFMPREGGTLIDMKMRIGKELKQAKRKNIIPLTYMNMWAIKALPEFQREEDV
SVSNAPGSCIIDCENHTRKKSOKETBELHCEYLAEPYMAOSTONVDNOLOEYVPEP
LKLBKSGELVYBESKRCGTSPLPACQVPTLPQVQYKENTQOPVAYQKMPBAL
QYRPPEQYQYQGMPPAPQGRAPYQPPPTRRILNTPAPRQKQSKLHEIIDKSKEDQ
TEAQPTFLTRPMPGGAQEGEPPTVEAYKFSPIKLDKMGKGVQYQPNSPYMT
LLDSQAFATLPIYDMEIILAKSSLSPSQFLQFTKWIIDQOEVRRRANPANIIDA
DOLIGIQNMSTIISQALMONEAIEYRAICLAWKEIIOBPGSTCFNTRVQVSE
YPPDPAIQDPAQKSIADERAKVIYELMAENANPCOGAIPRLKQVAGSDVSE
YVKQCDIGGAMERKAMLMQAQITGVVLGGVTRTGRKCTYCGQIHLKKNCPYLNKON
ITIDATTTGRRPDLCPRCCKKMAQCSKDKDNGQPLSGNENRGOPOAPQDTGAF
PIQFPVPGFGQGPPLSOVFGQISQLPDYNNCPPPQAAVQ"

gene
CDS
'
3051. . 5669
/gene="pol"
3051. . 5669
/gene="pol"
/codon_start=1
/product="pol protein"
/protein_id="CA76885.1"
/db_xref="GI:1185947"
/db_xref="GOA:O9MJR4"
/db_xref="SPTRMBL:O9MJR4"
/translation="MLTDLRANVANVIDPMBELQPLSPSPMIPKDWELIIDKDFE
TIPLAEOCEKFAFTIPAINNKEBATPQMKVLPQGMINSPTICQTFVGHALPVEK
FSDQIHCIDIDILCAETKDKLIDCQTPLOAEVANAAGLAISDKIQTSPFHYLGMQ
IENKRIQKIDKILDKTLINDFOKLDGIDINWIRPLTGIPITAMSNLSLILGSDNL
NSKMLTPREAKELIYBEKIQSQKQIRIDPLAPOLLITATASPGIITONTDVE
WSPLPHSTVKTFTLDQIATLIGQTLRIKICGNPDGIWPLETVEQVQAFINSQ
AMKIGAPVGIIDNHYPKTIQFGLKLTWMLPKTIRREPLEALTVFQDGSNGGA
YSPQKREKVIITPYQSAQRAELVAVITVLQDFQPIIISDAVVOGATDEVALIK
YSMDOINQLENLQOTVARKNPFYITTHRAHNLGPRLTKNEAQDLVSSALIKV
OELAHANVAAGLKKKEEDVTWQAKDITQHCQCOYLHPIQEGAGNPGCLPMLM
QMDTVHPSRGLSYVHTVDTYSHTIWTATQTBESISHYKHLSCFAVMGPBEKTK
TDNGVYSKRAQFKLSQMKISHTTGIYPSNQQAIVERNRTLKTQLVKQKGGDSK
ECTTPQOMLNALTYLNFILYRNQTTSAEHLTGKSNPHEKLTIMNDKNKME
IGKTIYMRGACVSPENOLPWIPTRHLPYNEPIRDAKSKTSAETETOSSTVDS
ODEBNGVRRPRDEVALIHQEGRAANLGTTKEADANYSIKSEHKGDTPRRYAACSLDD
CINGKSYAKSSCS"

gene
CDS
'
5386. . 7470
/gene="env"
5386. . 7470
/gene="env"
/codon_start=1
/product="env protein"
/protein_id="CA76886.1"
/db_xref="GI:1185948"
/db_xref="GOA:O9YNA5"
/db_xref="SPTRMBL:O9YNA5"
/translation="MQKKAPRRRRHRNRPALTHKNKMYTSEBOMLTPESTKAAEPT
MAQLKQLOLTKTLENTKVTQTPESMLNALKVSNVSPREPAGAAALTYTAAYRP
PPEPLIRAVTMMNDPTEVYVNDVSWVGPIDIDRPAPEEGBMININISIVHAPICL
GRAPCIMPAAONMLVEPVPVSPICRETYHNVSSQSLRPVNTLODPSYRSKFRKQ
GKPECKEIPKSKNTEVLYWEECVANSAVILONNEFPTIIDMAPRGQFYNGSGQDS
CPSAQVSAVNDLTLESLGIKHKIKKLOSFPWMEGKGIETPRKPIVSPVSGREHPL
WRUYTSHHIRIINGNOTLETBRDKRPYTIIDNLSLTPLQSCYKPRYMLVGNVYIK
PDSQTTTENCERLLTICIDSTFMQHRILLVAREGVIYVSMDEPWEASVHIITLEV
LKQVGNLRKRPFTLIVIMGLAVITAAVAVALLAQSSVSNFVNDWOKNSTRLN
SQSSIDQKALQINDLRQTVIWMGDRILSLEHFLQLCDMNNTSDFCITPOIYESELH
WMDVRHLQGEEDNLTLDISLKEQIPEASKAHLNLVGTREALAGVADGLANLPVW
VKITGSTIINLILIVLCFLLIYQCSQOOLARDSDHRRRAMTMAVLSKRGAGNVG
KSKKDQIYTVVS"

gene
CDS
'
7425. . 8300
/note="3' LTR"

ORIGIN

Query Match 100.0%; Score 150; DB 14; Length 8300;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGAGACCTCGGGCAGCAATAGCTGCTT 60
Db 7985 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGAGACCTCGGGCAGCAATAGCTGCTT 8044

Qy 61 GTAAGACCTGAGATGTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 8045 GTAAGACCTGAGATGTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 8104

Qy 121 ATTGCTATGATGCAAGACCTTTGCTTAC 150
Db 8105 ATTGCTATGATGCAAGACCTTTGCTTAC 8134

RESULT 9

AF490464

LOCUS AF490464 8708 bp DNA linear PRI 08-APR-2002
DEFINITION Homo sapiens HERV-K' long terminal repeat, complete sequence; and
envelope glycoprotein gene, partial cds.
AF490464
VERSION AF490464.1 GI:20067082

ACCESSION
KEYWORDS
SOURCE
ORGANISM

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 8708)
Reus, K., Stuhr, T., Mayer, J. and Meese, E. U.
Haplotype analysis of Human Endogenous Retroviruses: A genetic
variant of HERV-K(HML-2.HOM) with an intact YXDD motif of reverse
transcriptase

REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 8708)
Reus, K., Stuhr, T., Mayer, J. and Meese, E. U.
Direct Submission
Submitted (06-MAR-2002) Humangenetik, Universitaet des Saarlandes,
Kirtbergstrasse, Homburg 66421, Germany
Location/Qualifiers

FEATURES

source

1..8708
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/note="from Asian sample
endogenous virus: HERV-K'"
239..>8708
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/note="endogenous virus: HERV-K'; from Asian sample"
239..1206
1350..3350
/note="nonfunctional gag protein due to mutation"
3153..4155
/note="nonfunctional protease protein due to mutation"
4115..6986
/note="nonfunctional polymerase protein due to mutation in
gag"
6688..>8708
/codon_start=1
/product="envelope glycoprotein"
/protein_id="AA095525.1"
/db_xref="GI:20067083"
/translation="MNPSEMQKAPRRRRHRNRP.LTHKNKAVTSEBQKLPSTKK
AEPPTMAQLKLTQLATKYLENTVOTPSMILAALMIVSMVSLPMPAGAAANTY
YMAVPPPLIRATVMDNPLEVVDNSVWPGPIDRCAPKPEEGMNIISGYHY
PPICLGAPGCLMPAVONMLVEVTPSPICFTYHMGSLRRVAVYLDQFSYRSI.
KFRKGRPCREIPKESKNTSEVLVMECVANSVAILQNNERTIIDNAPROQFYHNS
GQIQCSCSAQVSPAYSDSDTESLDKHGKLOSFPYPMWSEKGIISTPRKIVSPG
EHEPEMLTAVASHHRIWISGNQLEETDRKPEFTIDNSLTVLOSVCVPEYMLVVG
NIYIKPSQITTCNCGLLTCIDSTPRWQRIILVPRABEGVITIVSDMPREASPSVH
ILTEVLKGVILNRSKRFIFTLAVMGLIATVAIVAGVALHSSVQSYNFPNDQKNS
TRINSSQSIDQKLANQINDLRQVIMWGRLMELERFOLQCDMTSDFCITQIYN
ESEHVMVRRHLQGRDNLDLIDSKLEQIFESKSKHLNLPVGTETIAGVADGAILN
NPVTWVXTISSTIINLILVLCIFCLLVCRCTQOLRRSDHREBAMMT"

CDS

LTR
misc_feature
misc_feature
misc_feature
CDS

ORIGIN

Query Match 100.0%; Score 150; DB 9; Length 8708;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGAGACCTCGGGCAGCAATAGCTGCTT 60
Db 799 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGAGACCTCGGGCAGCAATAGCTGCTT 858

Qy 61 GTAAGACCTGAGATGTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
Db 859 GTAAGACCTGAGATGTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 918

Qy 121 ATTGCTATGATGCAAGACCTTTGCTTAC 150
Db 919 ATTGCTATGATGCAAGACCTTTGCTTAC 948

RESULT 10

AY037929

LOCUS AY037929 9463 bp DNA linear VRL 29-OCT-2001
DEFINITION Human endogenous retrovirus K115 complete genome.
AY037929
VERSION AY037929.1 GI:16507983

KEYWORDS
SOURCE
ORGANISM

Human endogenous retrovirus K115
Human endogenous retrovirus K115
Viruses; Retroid viruses; Retroviridae.
1 (bases 1 to 9463)
Turner, G., Barbulessu, M., Su, M., Jensen-Seaman, M. I., Kidd, K. K. and
Lenz, J.
Insertional polymorphisms of full-length endogenous retroviruses in
humans
Curt. Biol. 11 (19), 1531-1535 (2001)

FEATURES

source

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (04-JUN-2001) Molecular Genetics, Albert Einstein College
of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA
Location/Qualifiers

1..9463
/organism="Human endogenous retrovirus K115"
/proviral
/mol_type="genomic DNA"
/specific_host="Homo sapiens"
/db_xref="taxon:166123"
1..960
1104..6740
/note="gag-pro-pol; two -1 frameshifts predicted to occur
to produce a fusion protein; the location of frameshifts
has not been determined"
6442..8541
/codon_start=1
/product="putative env"
/protein_id="AAL18259.1"
/db_xref="GI:16507984"
/translation="MNPSEMQKAPRRRRHRNRP.LTHKNKAVTSEBQKLPSTKK
AEPPTMAQLKLTQLATKYLENTVOTPSMILAALMIVSMVSLPMPAGAAANTY
YMAVPPPLIRATVMDNPLEVVDNSVWPGPIDRCAPKPEEGMNIISGYHY
PPICLGAPGCLMPAVONMLVEVTPSPICFTYHMGSLRRVAVYLDQFSYRSI.
KFRKGRPCREIPKESKNTSEVLVMECVANSVAILQNNERTIIDNAPROQFYHNS
GQIQCSCSAQVSPAYSDSDTESLDKHGKLOSFPYPMWSEKGIISTPRKIVSPG
EHEPEMLTAVASHHRIWISGNQLEETDRKPEFTIDNSLTVLOSVCVPEYMLVVG
NIYIKPSQITTCNCGLLTCIDSTPRWQRIILVPRABEGVITIVSDMPREASPSVH
ILTEVLKGVILNRSKRFIFTLAVMGLIATVAIVAGVALHSSVQSYNFPNDQKNS
TRINSSQSIDQKLANQINDLRQVIMWGRLMELERFOLQCDMTSDFCITQIYN
ESEHVMVRRHLQGRDNLDLIDSKLEQIFESKSKHLNLPVGTETIAGVADGAILN
NPVTWVXTISSTIINLILVLCIFCLLVCRCTQOLRRSDHREBAMMT"

CDS

LTR
misc_feature
misc_feature
CDS

GOTOSCPASQAVPAVDSLTRESIDKHKHKLQSPYPMWMEGKRISTPRKIVSPVSG
EHEBELRLTASHHRIIRISNGOTIETDRKRPFTTVDNSLITPLGSCVPPYMLVIG
NIVIKPDSQITTCENCLTICIDSTENWQHRILVLRBSVMIPLVSDRWESPSVH
ILTEVLKGVNRSKRIFLTIAYIMGLIATYAAVAGVALSHSVQSVNVNGQKNS
TRLNMOSSIDOKLANQINDLRQTVIMWGRILMSLEHRFLOCDMNTSDCTIPQIYN
DSHHMDVNRHLOQREDNLTDLSKKEQIFPASKAHNLVGTETAIAGVAGLANL
NPVTWKTIGSTTIIINLILVLCFLILVCRCTQOLRDSHREBRAMTMVAVLSKSK
GAVNGSKRDOITVSV"

ORIGIN LTR 8496.9463

Query Match 100.0%; Score 150; DB 14; Length 9463;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGGACCTGCGGCGACGACATCTGCTT 60
DB 553 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGGACCTGCGGCGACGACATCTGCTT 612

QY 61 GTAAGACCTGAGATGTTTATGTATGATATCTTAAAGACAGCACTTAATCTTTAC 120
DB 613 GTAAGACCTGAGATGTTTATGTATGATATCTTAAAGACAGCACTTAATCTTTAC 672

QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 673 ATTGCTATGATGCAAGACCTTTGTTTAC 702

RESULT 11
AF164614 9472 bp DNA linear PRI 03-SEP-1999
DEFINITION Homo sapiens endogenous retrovirus HERV-K108, complete sequence.
ACCESSION AF164614
VERSION AF164614.1 GI:5802820
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 9472)
Barbulescu, M., Turner, G., Seaman, M.I., Delnard, A.S., Kidd, K.K. and
Lenz, J.
Many human endogenous retrovirus K (HERV-K) proviruses are unique
to humans
Curt. Biol. 9 (16), 861-868 (1999)
JOURNAL MEDLINE 99400989
PUBMED 10469592
2 (bases 1 to 9472)
Barbulescu, M., Turner, G., Seaman, M.I., Delnard, A.S., Kidd, K.K. and
Lenz, J.
Direct Submission
Submitted (02-JUL-1999) Molecular Genetics, Albert Einstein College
of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

FEATURES
SOURCE
1..9472 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1..9472
/rpt_family="human endogenous retrovirus HERV-K108"
1..968
1112..6749
/gene="gag-pro-pol"
join(1112..3108,3108..3915,3915..6749)
/gene="gag-pro-pol"
/note="-1 frameshifts presumed to occur within the codons
for the last amino acids in the gag and pro open reading
frames"
/codon_start=1
/product="Gag-Pro-Pol protein"
/protein_id="AADS1797.1"
/db_xref="GI:5802821"
/translation="MGQTKSKIKSVASYSLFIKILKRGVAVSTNLIKLFQIIIEQ

FCWPEQGTIDLKKMKRIGKELKQKRGKNGIPLTWMDMALIKALEPOTEEDSV
SVSDAGSGCITDGNENTRRKSKQETSGHCEYAEVWMSOTPNVDNOCQETIYPT
LKIEGPELVEBSEKPRGTSPLPAGQPVTLQPKQVKNETQPPVATQVPPAE
QXRPPESSYVYGPMPAPQAGRAPYQPPRLNPLAPSKQSKLHEIIDSKRGED
TEWQPPVLEPMPPEGAQCEPPPEVPEAKYKFSIKLKDMEKQVAGNRPVYMT
LDSIHGRHLLIYDWEILAKSLSBSQFLOFTWMIIDGQEVRRNRPANPVNIDA
DOLLGIGWMTLSQOALMONEAIEQVRAICRAWKTIQDPSTCGSPFNVRGSGS
YDFVVARLDVNAQKSLADEKARVITYELMAYENANECQSAITPLKQVAGSDVSE
YVACDGIAGAHKAMLMQAITGVVAGQVRFGRKCTNCGQIGHKQKQVLANON
ITIQATTTGRBEPDLCPRCKGKMAVSQCRSKFNQAGPLSGNEQSGQPAPOQTA
PIQPFVPOGFOGQPLSQVFOGISOIPQYNNCPPOAAVOQVLCITIQAVSLPGEP
POKTPGVYGLPKGTGVLILGRSSLNKGVOHTSVBDSYGEIOLVSSSIPMSA
SPMRIDKMDPLIILDKCEFTIPLAEQCEKAFTIPIANNEPAPTRFQWKLPOOM
OEGVLDTGADVSIILNOMPKNPQKAVTGIVGTASEYQSTBEILHCLCPDNE
STVQPMITSIPLNLWGRDILQWGAETPAPBSPTSQIKMTWGTPEKQJGKED
GLIKPVEAKINDEBEIGNFCGAATVEEPKPLPTWKEKPVWVQMPPLPKLEL
HLIANLEKRGHIEPSPWNSPVFVIOKSGKMRMLTDLRANAVIOPMGPOGPI
SPAMIKDMDPLIILDKCEFTIPLAEQCEKAFTIPIANNEPAPTRFQWKLPOOM
INSPTICQFVGRALQVREKSPDVIITIDDIILCAAFKDKLIDCYFLOAEVANA
GLAIAIDKIQSTPFFHYLQMOIENRKLKQKTEIRDTLKTLDNFQKLGDIWIRPT
LGIPYAMSNLSEPIILGSDVLSNKRMLTEPAKEIKLVEBKIQSAQINRIDLPLGL
LAFATASPTGIIIONTDVENSFLSHSVKFTLVLDQATLIGQTRILIKLQND
PKIIVPLFREQVQAFINSKAWKIGLANFVGIIDNHYPETKIFQELKLTMTLPIKT
RREPENALTFVTDSSNGKAAVGPKEVYIKTPYQSAQBAELVAVTVLQDDPQIN
IISDAVWVQATRDVETALIKYSMDQNLQFLQOTYRKRREPPYITHIRANTNP
GPITKANEQDLVSSALIKAGELHALTHYNAAGLKNPDYTHKQAKDIYQCTQCOV
LHLPTQEAGVNPRLCQNALMOMDVTHVSPGRLSYVHVTVDTYSHITATQYGBST
SHVKKHLISCFAWGVPEKIKTDNGPYCSKAKQKLSQMKISHITGI PNASQGAIV
ERTNRILTQLVYKQEGDSKECTIPOMOLNALLYLNFINIYRNOITTSAEQHLIG
KNSPHEGKLIIMKKNKTBIEGKIVTGKVCVSGENQULPVMIPTRLIKYNEBI
RDAKSTSAETENOSSTVSDODENQDWRDEVAIHQEGRAANIGTTMEADAVYK
ISREHKDTPNPRYACSLDDCINGAKSPYACSSCS"
6451..8550
/gene="env"
/gene="env"
/codon_start=1
/product="envelope protein"
/protein_id="AAD51798.1"
/db_xref="GI:5802822"
/translation="KNPSEWOKAPPRRRRRHRNRPILTHKKNKQVTSSEOMKLPSTK
AABPTMAOLKLTOLATKYLENTKYQTPESMLALMIVSMVSLPMPGAAAYNT
YAAVYFPEPLIAYTMNDPTEVYVVDVSWPPIIDRCQAKREBEGMNTISIGHY
PILCLGRAPCLMPAVQNMVVEPVTSICRFYHWSGMSLEPRVAYLDDFYSQSL
KPRPKGKPPCKEIPKESKNTENVLWEECVANSVILLQNEFGTIIIMAPRGQFYNCS
GOTOSCPASQAVPAVDSLTRESIDKHKHKLQSPYPMWMEGKRISTPRKIVSPVSG
EHEBELRLTASHHRIIRISNGOTIETDRKRPFTTVDNSLITPLGSCVPPYMLVIG
NIVIKPDSQITTCENCLTICIDSTENWQHRILVLRBSVMIPLVSDRWESPSVH
ILTEVLKGVNRSKRIFLTIAYIMGLIATYAAVAGVALSHSVQSVNVNGQKNS
TRLNMOSSIDOKLANQINDLRQTVIMWGRILMSLEHRFLOCDMNTSDCTIPQIYN
DSHHMDVNRHLOQREDNLTDLSKKEQIFPASKAHNLVGTETAIAGVAGLANL
NPVTWKTIGSTTIIINLILVLCFLILVCRCTQOLRDSHREBRAMTMVAVLSKSK
GAVNGSKRDOITVSV"

ORIGIN LTR 8504.9472

Query Match 100.0%; Score 150; DB 9; Length 9472;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGGACCTGCGGCGACGACATCTGCTT 60
DB 561 GAGATAGGAAAAAACCCTTAGGCTGAGAGTGGGACCTGCGGCGACGACATCTGCTT 620

QY 61 GTAAGACCTGAGATGTTTATGTATGATATCTTAAAGACAGCACTTAATCTTTAC 120
DB 621 GTAAGACCTGAGATGTTTATGTATGATATCTTAAAGACAGCACTTAATCTTTAC 680

QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 681 ATTGCTATGATGCAAGACCTTTGTTTAC 710

RESULT 12
LOCUS HEN17832 10029 bp DNA linear VRL 21-OCT-1999
DEFINITION Human endogenous retrovirus K (HERV-K) elements, clone C7.
ACCESSION Y17832
VERSION Y17832.2 GI:4581240
KEYWORDS env gene; env protein; gag gene; gag protein; long terminal repeat; pol gene; pol protein.
SOURCE Human endogenous retrovirus K
ORGANISM Viruses; Retrovirdae; Retroviridae.
REFERENCE 1
AUTHORS Tonjes, R.R., Czauderna, F. and Kurth, R.
TITLE Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K
JOURNAL J. Virol. 73 (11), 9187-9195 (1999)
MEDLINE 99445825
FUBMED 10516026
REFERENCE 2
AUTHORS Tonjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1998) R.R. Tonjes, Paul-Ehrlich Inst., Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-61207 Langen, FRG
REMARK 3 (bases 1 to 10029)
REFERENCE 3
AUTHORS Tonjes, R.R.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) R.R. Tonjes, Paul-Ehrlich Inst., Paul-Ehrlich Str. 51-59, P.O. Box 1740, D-61207 Langen, FRG
COMMENT On Apr 12, 1999 this sequence version replaced gi:4185937. Related sequence Y10390.
FEATURES
source location/Qualifiers
1..10029
/organism="Human endogenous retrovirus K"
/proviral
/mol_type="genomic DNA"
/specific_host="human"
/db_xref="taxon:45617"
/clone="C7"
/clone_lib="PI (Genome Systems, St. Louis, MO, USA)"
/note="Assigned to human chromosome 7"
1..83
/repeat_family="Alu"
243..426
/rpt_family="Alu"
427..1394
/note="5' LTR"
1397..1414
1538..3538
/gene="gag"
1538..3538
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="CA76878.1"
/db_xref="GI:4185938"
/db_xref="GOA:O9YNA9"
/translacion="MGQTSKIKSKVSYSLFIKILKRGVKNVSTNLTKLFOITBQFCWPFQGTDLIDKMRIGKELKQAGRKNIIPVTWMAIIKALBEPQEDSEVSVDAFSCIIIDNENTRKSOKETEGLEHYEVPVMASTONVDVNOOEVIYPTLKLKGGKPELVGSESKRGTSPLPAGOVPTLOPOKOVKENTOPVAVQVPRATQYRPSPQYGRGMPAPGGRAPYPPPTRRNLTPPSPQSKHEITDKSKEDTEAMQPVLTLEPMPGSGAOGEPPTVEARVKSFSIKLDMKEGVKQYSPNRYMTDLDSIAHGRLIPYDWEILAKSLSPSQFLQPKTWIDGOVQVRRNRANPPNIDADOLIGCONSTISQALMONEALEIVRAICLAMEKIDPGSCSPFNRAQSGSEVYPPFARLOVAAKSIDAKARKIVIELMAYENNPCCOSAIKRLKGVAGVLSSEYVACDSIGAMVAMAMQALITGVIVGOVTRTGRCCVNCGOIGHKACPKLVNKNITIOATTTGRRPPLCPCKCKKGMASQCSKPKKCPISGNEORQOPAPOOTGAFPIQPFVQGFQGOOPUSOVFOGISOLPOTNCPPOAAVQO"
gene
4557..7175
/gene="pol"
4557..7175
CDS

/gene="pol"
/codon_start=1
/product="pol protein"
/protein_id="CA76878.1"
/db_xref="GI:4185939"
/db_xref="GOA:O9YNA9"
/translacion="MLTDLRAVNAVIOPMGLPGLSPAMIPKMWLIITDKCFETIPLAEDCEKFAFTIPAINNKEPATFOKVLPOGLNPTICOTVGAALQVREKPSDCVTHCIDDLICAEFTDKLIDCTPTQABVANAAGLAIASQIOTSPIHFGMOIBNRKIPQKIEIRKDTLKTUNDFOKLGSIDINIRPLUGIPTAMSLVSLRSDPLNSKMLTPPEATKEIKLVEEKIQSAQINRDIQBLIFATASPGIIGIQTNDLVEWSPLHSTVITFTLIDQIATLIGOTRLRIIKCGNDPKIYVPLTEQVROAFINSGAAMKIGLANPGLIIDNHYPKTKIQLKLTWILPKITRRRELEALTFVTDGSSNGKAAFYGKEXRVIKTPYOSARBAIVITVLDDPQINIIISDAVAVOATDVERALKYXMDDOANQFNLQCTVRGRNPFYTHIRATNLGPIPTKANEQADLVSSALITAOELHALTVNAAGLKNFDVTVYKQAKDIOGHCTCOVLHPTDAGVNPGLCPNALMOMDVTHVSPFGRSLSYHVTVDYSHFIWATCOTGBESTSHVKHLSCFVWGVPEKIKTDNPGYCSARFOFLSQMKISHTTGIPYNSOGAIVERTNRLTKTOLVWKEGDESECTTPQWQNLALYTNFLNINYNQTTTSAEOHLTGKNSPHEGLIMWDSKXKWEIGKVIITWGRGPACVSPENOLPWIPTRHKEVNEPIRDAKKSATETSOSSTVSDQENGVRRTRDERAIIQGBRANLITTKADAVSYKISREHKGDTTPREYAACSLDCLNGKSPYACRSCS"
6892..8976
/gene="env"
6892..8976
/gene="env"
/codon_start=1
/product="env protein"
/protein_id="CA76880.1"
/db_xref="GI:4185940"
/db_xref="GOA:O9YNA9"
/db_xref="SPTREMBL:O9YNA9"
/translacion="WQKAPRRRRRRNRAPLTHKNNKNTSEOMKLBSTKAEPTWAQDKLTQATLXYLTENTKVTQTPESMLALMIVSNVSLPMPAGAAANLYWAVYPFPLIRAVTMNDNPEVYVNDVSWVWGPIDIDBCPAPREBGMNINISYHVPICLGRAGCLMPAVQMWLVEVPVSPICRTYHVMGSRPRVNYLQDSYRSKLFRRKGRCPKBEIPKESKXTEVLVWBECAVANAIVLONNEFETIIDMAPRGQFYNSGQOTSCSAOVSRAVDSPLTESLIDGCHKKLSFVPMWGEKGIETPPRKIVSPSGPBPBLWRUTVASHIRINSNOTLETTRDKPFTTIDLNSLTVPLQSCVPPYMLVGNIVIKPDSQITTCENCRLLTCTDSTFNMQHRLILVAREGWI PVSMORPMWSPSVHILTEV LKGVNLNSKRFIFTLIAVINGLIAVATAVAGVALSSVQSVNFVNDQKNSRLNLSOSSIDQKLANQINDLRQTVIMMGDRMLSEHFLOCDNNTSDFCTPQIYNESEHMDWVRHLQGRENLTLDISKLEQIFEASKAHNLVPGREAIAGVADGANLPVWVKITIGSTTINILILVCLFCLILVCRCTQOLRDSDBRERAMTAVALSKRKGAVGKSKRDQIVTVSV"
8931..9898
/note="3' LTR"
9899..9903
/rpt_family="Alu"
ORIGIN
Query_Match 100.0%; Score 150; DB 14; Length 10029;
Best Local Similarity 100.0%; Pred. No. 1.4e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCCCTTTAGGCGCTGAGCGTGGACCTGGCGGCGACGATACTGCTTT 60
Dy |||||
Db 9491 GAGATAGGAAAAACCCCTTTAGGCGCTGAGCGTGGACCTGGCGGCGACGATACTGCTTT 9550
QY 61 GTAAGACCTGAGATGTTTATNGTATGATCTATAAGACAGCACTTAATCCTTAC 120
Dy |||||
Db 9551 GTAAGACCTGAGATGTTTATNGTATGATCTATAAGACAGCACTTAATCCTTAC 9610
QY 121 ATTGCTATGATGAAGAAGACCTTTGTTTAC 150
Dy |||||
Db 9611 ATTGCTATGATGAAGAAGACCTTTGTTTAC 9640
RESULT 13
LOCUS AF074086 19221 bp DNA linear PRI 11-MAR-2002
DEFINITION Homo sapiens tandemly repeated human endogenous retrovirus HERV-K

(HML-2.HOM), complete sequence.
ACCESSION AF074086
VERSION AF074086.2 GI:9558700
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 19221)
Meyer, J., Sauter, M., Racz, A., Scherer, D., Mueller-Lantsch, N. and Meese, E.
TITLE An almost-intact human endogenous retrovirus K on human chromosome 7
JOURNAL Nat. Genet. 21 (3), 257-258 (1999)
MEDLINE 99178257
PUBMED 10080172
REFERENCE 2 (bases 1 to 19221)
AUTHORS Reus, K., Mayer, J., Sauter, M., Scherer, D., Muller-Lantsch, N. and Meese, E.
TITLE Genomic organization of the human endogenous retrovirus HENV-K (HML-2.HOM) (ERV-K) on chromosome 7
JOURNAL Genomics 72 (3), 314-320 (2001)
MEDLINE 21295053
PUBMED 11401447
REFERENCE 3 (bases 1 to 19221)
AUTHORS Mayer, J., Mueller-Lantsch, N. and Meese, E. U.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1998) Institut fuer Humangenetik, Universitaet des Saarlandes, Homburg 66421, Germany
REFERENCE 4 (bases 1 to 19221)
AUTHORS Reus, K., Mayer, J. and Meese, E. U.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-2000) Institut fuer Humangenetik, Universitaet des Saarlandes, Homburg 66421, Germany
REMARK Sequence update by submitter
COMMENT On Jul 26, 2000 this sequence version replaced gi:4456987.
FEATURES
source
1. 19221
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/clone="Y20-B19"
/clone_1db="Lawrence Livermore National Laboratory Single Chromosome Library: L07NCO1"
1. 17976
/note="HML-2.HOM; tandemly repeated"
/rpt_family="Human endogenous retrovirus HENV-K"
/rpt_type="dispersed"
1. .968
971. .988
/note="lysine primer binding site"
1112. .3112
/gene="gag"
1112. .3112
/gene="gag"
/codon_start=1
/product="gag protein"
/protein_id="AAE88165.1"
/db_xref="GI:9558702"
/translation="MGQTSKIRSKVASYLSFIKILKRGVNVSTKNLIKLFQIIIEQ
FCMPFPEOGTLDLDMKRIKELKQAGKKNIIPLTYWDMALIIKALFPOTEBSV
SVSDPSSCIIDNENTRKKSOKEBELHCEYVAPEMAOSDNRNVDNLOAEVYIPET
LKLEGGKPELVGSESKRGTSPLPAGOVPTLQPOKOVKENTOPPVAAOYPAEEL
OYRPPESQGYRPMPPAPOGRAPYPOPPTRRLLPTAPPSRQSKLHEITDKSKED
TEAWQFVTLPEMPGGAQGEPTVEARKSFSLIKLDMKEGVAKQYGPNSPYMT
LPLSIAHGRLIPYDWEILAKSLSPQFLQFTKMWIDGOEVRNRANPPNID
DOLLIGONSTISQALQNEAIEQVRAICLAWEKIDPGSTPSFNTVRSKSEK
YVPEVARLOVAOSIADEKARKIVELAYENANPECOSAIKPLKGVAGSVISE
VYKACDGIAGAMKHAMLAOLINGVIGOVFTGGRKCVNCGOIKLKKKCPVLNKN
ITTOATTGRRPPLCPCKKKGKMAQCSKSKPKNCPISGNEORGOPOAPOOTGAF
PIPFVPOGFQGGPPLSVQFGISQLPYNCCPPPOAAVQ"
CDS
<2915. .3918

/note="translated by frameshift"
/codon_start=3
/product="protease"
/protein_id="AAE88166.1"
/db_xref="GI:9558703"
/translation="NATIVGKAKGASGPTTNWGISNAGICSSGSGSTTPPVPSV
GNKPVTTIQQLSPATSGSAVVDCTIQASVILGEPQKTPGTVGPPPLKGVGLIG
RSILNKGQIHITSVSDSDYKGEIQLVSSIPWSAPDRDIAQLLLPLVIGKNSGI
KRIGIGSTDPGKAIVASQVSENPVCAIIQKQFEGVLVDGTGADVITIALNWPK
NMPKOKAVGLVIGTASVSEVOSTEILHCLGPDNOSTVQPMITSIPLNMGRDLQO
KGAETMPAPSVSPISQKIMTMGYIRGGLGKNDEGIIPIVEAKINOREGICNFC"
CDS
<3878. .6749
/gene="pol"
/gene="pol"
/gene="pol"
/note="translated by frameshift"
/codon_start=2
/product="polymerase"
/protein_id="AAE88167.1"
/db_xref="GI:9558704"
/translation="NKSRRRRRESILGATVEPPKPIPLTWKTEKPVWVNCPLPQ
KLEALHLANEQLEKHIERSFSPNSPVFVIOKSGKRWMLTDLRANAVIQMPEPL
OGLPSPAMIPKMDWPLIITLDNDCPTPLADQCKKPAFTPIAINKNBAPRFOKY
LROGMNSTIQCOTFVGRALOYRREKSPSCIIHCIDDLICAAETDKLIDCTPQO
EVANAGLAIASDKITQSTPFIYQWQIENRKIKPKQIEIKQTLKLNDAQKLDGIN
WIRPTGIPTYMSNLFSILRGDSDLNSRMLPEATYKIKLVEEKIQSAQINRIDL
APLOLIPATYMSPTGIIIONTDLVMSPLSTVPTVTLVLIQOITLIGQTRLIK
LCGNDDKIVPLETKEOVQAFNSGAMKIGLANFVGIIDNHYPKIKIROPILKLTWII
LKPTRREPLENALATVFTQSSNGKAAVYGPERRYKTPYQASORAEVLAVITLDFE
DOPINISDAVVOATRDVETRALIKYSDDOLNQLPNLQOYVRKANPEFYTHIRA
HTNLPGPLTQBNQDLVLSALIKQELHALHVNAAGLKNKFDVTWKAKOIVQO
TQOVVHLPTQAGVNPGLCPVALMOQVTHVPSGVAVVAVDTVSHFIMATQ
TGESSTHVKKHLISCPAVWVEPKIKTDNGPQCSAPQKFLSQMKISHTTGIPVNSO
GQAIIVTRNLTQLOKQEGDESKCTTPQOMQLALYTLNPLNIYRQUTTSASQ
HLTKKNSPHEGLIMWKKNKTKTIGVITMGREFACVSPENDELPMVIPTRHKE
YNEPITDAKTSASERTSQSTVBSQDNQDNQDRRTDVAIQHQBGRANLGTTKAD
AVSYKTSREHKGNTREYAACSLDDCINGKSPVACRSCS"
CDS
6451. .8550
/gene="env"
/gene="env"
/codon_start=1
/product="envelope"
/protein_id="AAE88168.1"
/db_xref="GI:9558705"
/translation="NMPSEMOKAPPPRRRRHRRARPLTHKNKMTSEBOMKPTSK
ABEPTTAOLKKTTLQATKYLENKVTQTPESMLALMTVSVVSLPMPGAAASVY
YAAVYPPPLIAVLTMDNPTVYVNDSDVWVEPIDRPAKREBEGMNIISGYHT
PPLCIGRAPGCLMPAVQNLVEVPTVSPICRFYTHAVSGMSLTPRNVYLDSEYQSL
KRRPKGCPQPKIPKSKNTEVLWVECVANSVAILQNEFGTIDMAPKIGQFYNCS
GOTOSCPASQVSPAVSDLTSLDKIKKKLOSFTYMEGEEKISTPRPKIVPVGSP
EHEMLRLVASHHRIIRISGNQLETFRDKRPFYITDLSLSLTVPLQSCVPPMLVVG
NIVIKRDSQITICENCRLICTIDSTFNWQHRIILVARGSWATPVAMDREWESPVH
ILTEVUKGVANSKRIFITLIVAIMELIANTATAVAGVALHSVOSVNDVNMOKRS
TRLNNSQSIDQLANQINDLRQTVIMWQGRLSBHRFQLOQDMNTSDCTIPQIYN
ESEHMDWVRHLQGRBDNLTDISLKKQOIFEASVAHLNLVGTALAGVAGLANL
NPVTWTKTIGSTTILNLIILVLCFLLLVCRCTQLRDRSDHREBAMTMVAISRRK
GGVNGSKSRDOITVSV"
CDS
8505. .9492
9475. .9492
/note="lysine primer binding site"
9578. .9586
/note="splice donor site"
9616. .11616
/gene="gag"
9616. .11616
/gene="gag"
/codon_start=1
/product="gag"
/protein_id="AAD21095.2"
/db_xref="GI:9558701"
/translation="MGQTSKIRSKVASYLSFIKILKRGVNVSTKNLIKLFQIIIEQ
FCMPFPEOGTLDLDMKRIKELKQAGKKNIIPLTYWDMALIIKALFPOTEBSV

```

SVSDAPGSCIIIDCNENTRKKSOKEBGLHCEYVAEYMAQSTONVDYNOLOEVITYPRT
LKEGKPELVGSESKPRGTSPLPAGOVPTLQPOKOVENKTPQPVAYQVAPAEI
QVRPEPSOYXGPMGPAPGRAPYPOPPVRRLNPTAPSPKSGKHEIIDKSGED
TEAMQFVTLPEMPMREGQAQEGEPPTAEATYSIKLMDKMGUQXQYSPMYMT
LDSIANGHRLIPDWEITLAKSSISPSQFLOFTKWIIDGOEVRNRANAPPNITDA
DOLIGQNNSTISQALMONEALEQVRAICLRMEKIIDQGSTCPSFNVRQSKSK
YPRFVALQDVAOKSISADEKARKIVELMAYENANPCQSAIKELKGVAGSPVISE
YVACDIDIGAMKAMAMADAITGVVLGVCYRTGRACYNQIGHLKKNCPVANKON
ITTOATTGREGPDLCPRCRKKMKHMAOCSSKPKPNOPISGNRQROQPOAPQGTAF
PIQFVQROGQOCPPLSVQFGISQPLRYNNCPPPAAVQO"
<11419..12422
/translate="translated by frameashift"
/codon_start=3
/product="protease"
/protein_id="AAD1096.1"
/db_xref="GI:4456989"
/translation="WATTVGRKAKGAPSGPTTMMKINSATSCSGSEGTTPVPSVS
GNRPVTTIQQLSPATSGSAVVDLCTIQAVSLPEEPQKPTGVYGLPKGTVGLIG
RSSILNKGVQIHITSVDYKGEIQIYSSISPMASPRRIQALLPLPIKGNSEI
KRIGLGSTPTGKAAVMAQOVSENRVPCAIIQKQEGVDTGADVSIILNQMPK
NMPKOKAVTGVNGIGTASVYQSTELHICGPDQESTVQPMITSIPLNMGRLDQO
WGAETMPASYSYSTOKIMTKNGIIRGKLGKNEBDIKIPEAKINQEREGINPC"
<12382..15253
/gene="pol1"
<12382..15253
/translate="translated by frameashift"
/codon_start=2
/product="polymerase"
/protein_id="AAD1097.1"
/db_xref="GI:4456990"
/translation="NKSRRKRNRSLGAATVEBPRIPLTWKTEKPEVMVNOPLPKO
KLEALHLLANOLEKHLIPEPSFSPMSPVLTQKSKKMTLDLRVNVAVMGEL
QPGIPLSAMPKMPLIITIDKCFPTIPLAEQCEKPAITPINKEPRTOMKY
LPGMLNSPTICQTFVGRALQPVNEKSDCTIHCIDIDILCAATPKKIDCTYFLA
EVANAGAIASIDKIQSTPRLYLQMDLNRKIKQKEIRIDTLKTNDOQLKNDIN
WIRPLGIPYAMSNLPSILRGSDLSNKKETPEAKIKLVEKIOSQINRIDL
APQLIFATASPTGIIIONTDLVEMSPHSVVFETLYLODIALTIGOTRLIK
LCNDPDKIVPLPKEQVROAFINSQWKIGLNFVGIIDNHVKTIKIPQIKLTMV
LPKTRREPENALTFTDSSNGKAAVTGPKXEVITTPYQSAQRAALVAVITLQDP
DQPNITISDAVYVQATRDVETALIKISMDQLQLFNLLQOTVRKKNFPYITTHIA
HTNLPGPLTANBOADLVSALIKAOELHALTHVNAAGLNKEDVWMAKADIVQHC
TQCCVHLPTQEGVNPRLCPNALMQMDVTVPSFRLSYHVTVDTYSHFTYVNSQ
TGSTSHVKKHGLSCFPAWGVPEKIKTDNGVGSKAFQKLSOMKISHTTGIPYNSQ
GQALVERTNNTIKTQVLKQKQKSGKCTKPMOALATYANTANLYRQCTTSACQ
HLTGKRSPEHCKLIWKNKKNKTKWEGKITTMGRGCAISPGENQIPWIPTRHLKF
YNBPTRDKAKSTSAETQSSTVDSODEQNDVARDVAILHQBGAANLGTTEAD
AVSYKISREHKGDTPREYVACSLDDICINGKSPYACRSSCS"
14955..17054
/gene="env"
14955..17054
/translate="env"
14955..17054
/translate="env"

Query Match      100.0%; Score 150; DB 9; Length 19221;
Best Local Similarity 100.0%; Pred. No. 1.3e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

LOCUS AC072054 154037 bp DNA linear PRI 08-OCT-2003

```

```

DEFINITION Homo sapiens BAC clone RP11-33P21 from 7, complete sequence.
ACCESSION AC072054
VERSION AC072054.10
KEYWORDS GI:18056706
SOURCE HGC.
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 154037)
Sulston, J.E. and Wilson, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
MEDLINE
PUBMED
9847074
2 (bases 1 to 154037)
Du, H., Edwards, J., Haekenson, W. and Spalding, L.
The sequence of Homo sapiens BAC clone RP11-33P21
Unpublished (2001)
3 (bases 1 to 154037)
Waterston, R.H.
Direct Submission
Submitted (07-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 154037)
Waterston, R.H.
Direct Submission
Submitted (04-JUN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 154037)
Waterston, R.
Direct Submission
Submitted (30-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 154037)
Waterston, R.
Direct Submission
Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 154037)
Waterston, R.
Direct Submission
Submitted (26-APR-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
8 (bases 1 to 154037)
Wilson, R.
Direct Submission
Submitted (08-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 4, 2002 this sequence version replaced gi:13786481.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0033P21
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

```

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCL11 human BAC library was made from the blood of one male donor, as described by Ossegeawa, K., Moon, P.Y., Zhao, B., Frangsen, B., Tatenio, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pictet de Jong and coworkers at <http://www.chori.org>

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-552B11 the clone sequenced to the right is RP11-32P3. 2000 bp overlap. Actual start of this clone is at base position 22189 of RP13-552B11 actual end is at base position 2000 of RP11-32P3.

Location/Qualifiers

1. 154037

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="7"

/map="7"

/clone="RP11-33P21"

/clone_1lb="RPCL11"

1. 550

/rpt_family="L1"

/rpt_family="ERV.L"

4289. 4432

/rpt_family="MIR"

4964. 5214

/rpt_family="Alu"

5233. 5483

/rpt_family="L1"

5500. 5520

/rpt_family="AT_rich"

5994. 6283

/rpt_family="Alu"

6337. 6661

/rpt_family="Alu"

6991. 7114

/rpt_family="L1"

7401. 7634

/rpt_family="Alu"

7641. 7759

/rpt_family="L2"

7797. 7843

/rpt_family="L1"

7964. 8144

/rpt_family="MaLR"

8148. 8458

/rpt_family="Alu"

8461. 8747

/rpt_family="Alu"

8755. 8806 (GGGA)n

/rpt_family=""

9349. 9499

/rpt_family="L1"

9513. 9762

/rpt_family="Alu"

9768. 9936

/rpt_family="Alu"

9938. 10243

/rpt_family="Alu"

repeat_region

10278. 10576

/rpt_family="L1"

repeat_region

10577. 10897

/rpt_family="Alu"

repeat_region

10898. 11020

/rpt_family="L1"

repeat_region

11187. 11411

/rpt_family="L1"

repeat_region

11412. 11711

/rpt_family="Alu"

repeat_region

11712. 11961

/rpt_family="L1"

repeat_region

11973. 12264

/rpt_family="Alu"

repeat_region

12265. 12610

/rpt_family="L1"

repeat_region

12611. 12907

/rpt_family="Alu"

repeat_region

12908. 13739

/rpt_family="L1"

repeat_region

13743. 13876

/rpt_family="L1"

repeat_region

13885. 13934

/rpt_family="MER1_type"

repeat_region

14077. 14139

/rpt_family="MIR"

repeat_region

14576. 14863

/rpt_family="Alu"

repeat_region

15079. 15379

/rpt_family="Alu"

repeat_region

15648. 16009

/rpt_family="Alu"

repeat_region

16015. 16322

/rpt_family="Alu"

repeat_region

16326. 16454

/rpt_family="Alu"

repeat_region

17256. 17384

/rpt_family="L1"

repeat_region

17389. 17628

/rpt_family="Alu"

repeat_region

17639. 18045

/rpt_family="ERV.L"

repeat_region

18046. 18454

/rpt_family=" (TA)n"

repeat_region

18455. 18515

/rpt_family="ERV.L"

repeat_region

18516. 18576

/rpt_family="Alu"

repeat_region

18643. 18952

/rpt_family="Alu"

repeat_region

19052. 19447

/rpt_family="MaLR"

Query Match 100.0%; Score 150; DB 9; Length 154037;
Best Local Similarity 100.0%; Pred. No. 1.2e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAGATAGGGAAAAACCGCCTTAGGCGCTGAGAGTGGGACCTGGGGGAGCAATCTGCTTT	60
DB	46857	GAGATAGGGAAAAACCGCCTTAGGCGCTGAGAGTGGGACCTGGGGGAGCAATCTGCTTT	46798
QY	61	GTAAGCAGTGAATGTTATGTGTATGCAATCTAAAGCAGACGACTTAATCCTTTAC	120
DB	46797	GTAAGCAGTGAATGTTATGTGTATGCAATCTAAAGCAGACGACTTAATCCTTTAC	46738
QY	121	ATTGCTATGATGCAAGACCTTTGTTAC	150
DB	46737	ATTGCTATGATGCAAGACCTTTGTTAC	46708

RESULT 15
AC134684/c

Job time : 629.802 secs

```
repeat_region /rpt_family="L1MB5"
complement(14879..15054)
repeat_region /rpt_family="L1PA8"
15180..15206
repeat_region /rpt_family="AT_rich"
16055..16134
repeat_region /rpt_family="MADE1"
16711..16783
repeat_region /rpt_family="AT_rich"
17006..17334
repeat_region /rpt_family="AluSp"
17434..18204
repeat_region /rpt_family="L1MBd"
18205..18834
repeat_region /rpt_family="L1MBd"
18835..19070
repeat_region /rpt_family="L1MBd"
complement(24651..24944)
repeat_region /rpt_family="AluSq"
complement(26607..26682)
repeat_region /rpt_family="MIR"
27183..27520
repeat_region /rpt_family="L1MC5"
complement(27584..27860)
repeat_region /rpt_family="L1MC5"
28775..29075
repeat_region /rpt_family="L2"
29234..29336
repeat_region /rpt_family="L2"
29353..29466
repeat_region /rpt_family="L2"
complement(30426..30581)
repeat_region /rpt_family="MIR3"
complement(30729..30926)
repeat_region /rpt_family="MER4B"
complement(30927..31245)
repeat_region /rpt_family="AluSx"
complement(31246..31293)
repeat_region /rpt_family="MER4B"
complement(31337..31708)
repeat_region /rpt_family="MER4B"
32262..32447
repeat_region /rpt_family="MIR3"
33391..33486
repeat_region /rpt_family="MIR3"
34024..34052
repeat_region /rpt_family="AT_rich"
complement(34310..34400)
repeat_region /rpt_family="LTR16B"
34401..34713
repeat_region /rpt_family="AluJo"
complement(34714..34881)
repeat_region /rpt_family="LTR16B"
35338..35931
```

Query Match 100.0%; Score 150; DB 9; Length 175202;
Best Local Similarity 100.0%; Pred. No. 1.2e-35;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAGATAGGGAAGAAACCGCCTTAGGCGCTGGAGTGGGACCTTCGGGCGAGCAATACTGCTTT 60
DB 64860 GAGATAGGGAAGAAACCGCCTTAGGCGCTGGAGTGGGACCTTCGGGCGAGCAATACTGCTTT 64801
QY 61 GTAAAGCAGTGAAGTGTGTTATGTCATATGCAATCTAAAGCAGCACTTAATCCTTTAC 120
DB 64800 GTAAAGCAGTGAAGTGTGTTATGTCATATGCAATCTAAAGCAGCACTTAATCCTTTAC 64741
QY 121 ATTGCTATGATGCAAGACCTTTGTTAC 150
DB 64740 ATTGCTATGATGCAAGACCTTTGTTAC 64711
```

Search completed: February 27, 2004, 23:57:24


```

1      CITY: San Francisco
2      STATE: CA
3      COUNTRY: USA
4      ZIP: 94111-3834
5
6      COMPUTER READABLE FORM:
7      MEDIUM TYPE: floppy disk
8      COMPUTER: IBM PC compatible
9      OPERATING SYSTEM: PC-DOS/MS-DOS
10     SOFTWARE: Patent Release #1.0, Version0.00000000
11     CURRENT APPLICATION DATA:
12     APPLICATION NUMBER: US/08/724,394A
13     FILING DATE: 01-OCT-1996
14
15     CLASSIFICATION: 536
16
17     ATTORNEY/AGENT INFORMATION:
18     NAME: Flits, Renee A.
19     REGISTRATION NUMBER: 35,136
20     REFERENCE/DOCKET NUMBER: 017957-000100
21     TELECOMMUNICATION INFORMATION:
22     TELEPHONE: 415-576-0200
23     TELEFAX: 415-576-0100
24
25     INFORMATION FOR SEQ ID NO: 22:
26     SEQUENCE CHARACTERISTICS:
27     LENGTH: 246240 base pairs
28     TYPE: nucleic acid
29     STRANDEDNESS: not relevant
30     TOPOLOGY: not relevant
31
32     MOLECULE TYPE: cdna
33
34     FEATURE:
35     NAME/KEY: misc feature
36     LOCATION: 1..246240
37     OTHER INFORMATION: /note= "H1A-H. CONTIG"
38
39     OS-08-724-394A-22

```

Query Match	81.6%	Score 717.4	DB 2	Length	246240
Best Local Similarity	91.4%	Pred. No. 5.9e-244			
Matches 805	Conservative	0	Mismatches	71	Indels 5
					Gaps 4

OY		1	GGTGGGAAAAAGCAAGAGATCAGATTGTCTGTGTCGTGTGAAGAAAGTACACT	60
Dd		104446	TGTGGGAAAAGAGAAGAGATCACTTTTACTGTGTCTGTGTAGAAAGAGTACACT	104505
OY		61	AGGAGACTCCATTTTGTATGTACTAAGAAAAATCTTGCCCTTGAGATTCTGTAATC	120
Dd		104506	AGGAGACTCCATTTTGTCTGTACTAAGAAAAATCTTGCCCTTGAGATTCTGTAATC	104565
OY		121	TATGACCTTAAACCCCAACCCCGTCTCTGTAAACAATGCTGTGTCCAATCAGGTTAA	180
Dd		104566	TA--ACCCTAGCCCAACCCGTGTGCTCTGTAGACATAATGTGTGTCAATCAGGTTAA	104623
OY		181	ATGCATTTAAGGGCGGTGCAGATGTGCTTTGTTAAACAATGTTGAAGCAGCATGCTC	240
Dd		104624	ATGCATTTAAGGGCTGTGCAGATGTGCTTTGTTAAACAATGTTGAAGCAGCATGCTC	104683
OY		241	CTTAGAGTCATCACACATCCCTTAATCTCAAGTACCAGAGGACACAAA-ACTAGCGAAG	299
Dd		104684	GTTAAGATCATCTCACTCCCTTAATCTCAAGTACTCAGGAGACAAAACACTGAGAAAG	104743
OY		300	GCCGACAGGACCTCTGCCCTAGGAAAGCCAGTATGTCCAGCGTTTCTCCCATGTGATA	359
Dd		104744	GCCACAGGGA CCTCTGCCCTAGGAAAGCCAGTATGTCTCAAGTTTCTCCCATGTGATA	104803
OY		360	GCTTGAAATATGCGCTTCGTGGAAAGGAAAGACTGACCGTCCCGACACCCGT	419
Dd		104804	GTCGAAATGTGCGCTTCGTGGAAAGGAAAGACTGACCGTCCCGACACCCGT	104863
OY		420	AAAGGCTCTGTGCTGAGAGAGATTAGTAAAGAAGAAAGATGCTTC-TTGCA GTTGAGA	478
Dd		104864	AAAGGCTCTGTGCTGAGAGAGATTAGTAAAGAAGAAAGATGCTTCAGTTGAGAG	104923
OY		479	CAGAAGAAAGGATCTGTCTCTGCCCTGACCCGCGGCANTGGAATGCTCGATATAAAC	538
Dd		104924	CAGAAGAAAGGATCTGTCTCTGCCCTGACCCGCGGCANTGGAATGCTCGATATAAAC	104983

QY	539	CCGATTGATATCTCATCTTACCTGAGATAGGGGAAAAACGCCCTTAGGGCTGGAGAGTGGAGC	598
Db	104984	CCGATTGATATTCATCTTACCTGAGATAGGGGAAAACTGCCCTTAGGGCTGGAGAGTGGAGC	1050434
QY	599	CTGGGGGAGCAATATTCGCTTTGTATTAACACTGAGATATTTATATGTATGCAATATCTTAA	658
Db	105044	ATGCTGGAGGAAATACGCTCTTCAAGTCATTGAGATGTTATATGTATGCAATATCTTAA	1051020
QY	659	AGCAGACGACTTAAATCCTTTACATTTGTCATGATGCAAAAGACCTTGTTCACATGTTGT	718
Db	105103	AGCAGACGACTTAAATCTTTTACCTTTGTTATGATGTCAGAGACCTTGTTCACGTTGTTAC	1051622
QY	719	CTGCTGACCCCTCTCCCACTATTCGTTGTGTATACCTGACACTCCGCCCTTCGAGAAAC	778
Db	105163	CTGCTGACCTTCTCTCCACATATATCTTTGACCTGTGCACATCCCTCTTCGAGAAAC	1052222
QY	779	ACCCACAGATGATGATTAATCTAATAAGGAACTCAGAGCGTGGCGGATCTTCATATGC	838
Db	105223	ACCCCAATATGATCAATTAATATCTAAGGAACTCAGAGCGTGGGATCTTCGATATGC	1052822
QY	839	TGAACGCTGTTCCCGCGGTCCCTCTCTTTCTCTATATA	879
Db	105283	TGAACACGCGTCCCTGAGCCCTTTTCTTTCTCTATATA	1053123

```

US-09-482-273-20/c
RESULT 5
Sequence 20, Application US/09482273
Patent No. 6534631
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 71 Human Secreted Proteins
FILE REFERENCE: P2030P1
CURRENT APPLICATION NUMBER: US/09/482,273
CURRENT FILING DATE: 2000-01-13
EARLIER APPLICATION NUMBER: PCT/US99/15849
EARLIER FILING DATE: 1999-07-14
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/092,922
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/092,956
EARLIER FILING DATE: 1998-07-15
NUMBER OF SEQ ID NOS: 267
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
LENGTH: 2525
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (5)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (10)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1354)
OTHER INFORMATION: n equals a,t,g, or c
US-09-482-273-20

```

Query Match	78.7%;	Score 692.2;	DB 4;	Length 2525;
Best Local Similarity	89.8%;	Pred. No. 2.4e-236;		
Matches 780; Conservative	17;	Mismatches 35;	Indels 37;	Gaps 4;

QY 14 AAGAGATATGATTTGTTACTGTCGTCGTGTAGAAAGATGACATGAGAGATCCCAATT 73
 Dd 1725 AAGAGATCGATTTGTTACTGTCGTCGTGTAGAAAGATGACATGAGAGATCCCAATT 1666
 QY 74 TTGTTATGTACTAAGAAAAATTCCTGCGCTTGAGATTCTGTAACTATGACCTTACC 133

1665 TTGTTTCG-----NATCTTAACCTTACCC 1640
134 CCAACCCCGTCTCTGTAACATGTGCTGTCACTCAGGTTAAATGATTAAGGCG 193
1339 CCAACCCCGTCTCTGTAACATGTGCTGTCACTCAGGTTAAATGATTAAGGCG 1580
194 GGTGAGAGATGTGTTGTTAAAGATGCTTGAAGGAGATGCTCTTAAAGTCATC 253
1579 GGTGAGAGATGTGTTGTTAAAGATGCTTGAAGGAGATGCTCTTAAAGTCATC 1520
254 ACCACTCCCTTAATCTCAAGTACCCAGGACACAAA-ACGCGAAGGCGCCAGGACCT 312
1519 ACCACTCCCTTAATCTCAAGTACCCAGGACACAAA-ACGCGAAGGCGCCAGGACCT 1460
313 CTGCTTGAAGAGCAGGATTTGTCACGTTTCCCATGATGATGATGATGATGATG 372
1459 CTGCTTGAAGAGCAGGATTTGTCACGTTTCCCATGATGATGATGATGATGATG 1400
373 CCTGCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 431
1399 CCTGCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1340
432 CTGAGAGGATTTAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGCA 491
1339 CTGAGAGGATTTAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGCA 1280
492 TCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 551
1279 TCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1220
552 CCATCTACTGAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 611
1219 CCATCTACTGAGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1160
612 TACTGCTTTGTAAG 671
1159 TACTGCTTTGTAAG 1100
672 ATCTTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 731
1099 ATCTTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1040
732 CCCCACATTTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 791
1039 CTCACATCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 980
792 CAGTAATCTAG 850
979 CAAATAATCTAG 920
851 CCCCCGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 879
919 CCCCCGGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 891

RESULT 6
US-08-991-789A-141
Sequence 141, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Prud'homme, Tony N.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 622-6031
INFORMATION FOR SEQ ID NO: 141:
SEQUENCE CHARACTERISTICS:
LENGTH: 9388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 141:
US-08-991-789A-141

Query Match 56.3%; Score 494.8; DB 3; Length 9388;
Best Local Similarity 79.1%; Pred. No. 2.1e-165;
Matches 709; Conservative 0; Mismatches 132; Indels 55; Gaps 8;
20 GATCAGATTGTTACTGCTGTCTGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 79
48 GATCAGAGTGTATGCTGTCTGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 107
80 TGTACTAAGAAAAATTTCTTCTGCTTGTGATTTCTTTAATGATGATGATGATGATGAT 139
108 TGTACTAAGAAAAATTTCTTCTGCTTGTGATTTCTTTAATGATGATGATGATGATGAT 167
140 CCGTCTCTGAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 199
168 CTGCTCTCAG 221
200 GATGCTGCTTGTGTAACAGATGCTTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 259
222 ---TATGCTTTGTTAAAG 278
260 CCTTAATCTCAAGTACCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 318
279 CTCTAATCTCAAGTACCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 338
319 AGGAAAGCAGGATTTGTCACAGCTTCTCCCAATGATGATGATGATGATGATGATGATG 378
339 AGGAAAGCAGGATTTGTCACAGCTTCTCCCAATGATGATGATGATGATGATGATGATG 398
379 GGAAG 408
399 GGAAG 458
409 CCGAAG 468
459 CCGAAG 515
469 GCAAGTGAAG 528
516 GCAAGTGAAG 575
529 GGTATTAAG 588
576 GGTATTAAG 635
589 GAGGTGAG 648
636 GAGGTGAG 695

OY	649	CATATCTAAACACAGCACTTAATCCCTTACCTGTATGATGAGAAACCTTGTTC	708
Db	656	CACATC--AAGCACACGACACT-TTCTTTAACTTATTATGACACAGAACCTTGTTC	752
OY	709	ACATGTTTGTCTGTCGACCCCTCTCCCCACAAATGTCTTGACCTTGACATCCCCCTC	768
Db	753	AC-GTTTTCCTGTCGACCCCTCTCCCACTATTACCTATTGGCTGCGCACATCCCCCTC	810
OY	769	TTTCGAGAAACCCACAGATGATCAGTAATACTAAGGGAACTCAAGAGCTGGCG----	823
Db	811	TCCGAGATGCTGAGATTAATGATCAATTAATACTGAGGGAACTCAAGAGCAATGTCCCT	870
OY	824	--GGATTCCTCCATATGCTGAACGCTGATGCCCGGGGCCCTCTTTCTTTCTCTA	877
Db	871	GTAGGTCCTCCGCTGCTGCTAGCGCCGAGTCCCTTTGGGCTCACTTTTCTTTCTATA	926

RESULT 7

US-09-062-451-141
Sequence 141, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT OF INFECTIONS: TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 141:
SEQUENCE CHARACTERISTICS:
LENGTH: 9388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-062-451-141

Query	Match	Similarity	Score	DB	Length
Best	Local	Similarity	79.1%	Pred. No. 2.1e-165	
Matches	709	Conservative	0	Mismatches 132	Indels 55; Gaps 8

Qy	20	GATCAGATTGTA	CTGTGTCGTGTAGAAAGTAGACATGAGAGACTCCATTTTGT	TA	79
Db	48	GATCAGACTGTGTA <th>CTGTGTCGTGTAGAAAGTAGACATGAGAGACTCCATTTTGT</th> <td>TA</td> <td>107</td>	CTGTGTCGTGTAGAAAGTAGACATGAGAGACTCCATTTTGT	TA	107
Qy	80	TGTACTAAGAAAAT <th>CTTCTGCGCTTGAGATCTGTGTAATCATATACCTTACCCCAAC</th> <td>CC</td> <td>139</td>	CTTCTGCGCTTGAGATCTGTGTAATCATATACCTTACCCCAAC	CC	139
Db	108	TGTACTAAGAAAAT <th>CTTCTGCGCTTGAGATCTGTGTAATCATATACCTTACCCCAAC</th> <td>CC</td> <td>167</td>	CTTCTGCGCTTGAGATCTGTGTAATCATATACCTTACCCCAAC	CC	167
Qy	140	CCGTGCTCTCTG <th>AAACATGTGCTGTGTCCACTCAGGGTTAAATGAGTTAAGGGCGGTGCA</th> <td>TA</td> <td>199</td>	AAACATGTGCTGTGTCCACTCAGGGTTAAATGAGTTAAGGGCGGTGCA	TA	199

```

Db      168 CTGTCCTCACAGACAGCATGTGCTGTGTGACTCAAGATTCATGATTAAGGCG----- 221
Qy      200 GGATGTGCTTTTGTTTAAACAGATGCTTTGAAGGACGATGCTCTTTAAGATCAACACACT 259
Db      222 ---TATGCTTTGTAAAAAAGCTGTGAAGATAAATATGCTTTTAAAGTCATCACCAT 278
Qy      260 CCTAATCTCAAGTACCCAGGGACACAA-AACTCGCGAAGGCCGACGGACTCTGACT 318
Db      279 CTCTAATCTCAAGTACCCAGGGACACAACTACTGCGAAGGCCGACGGACTCTGTCT 338
Qy      319 AGAAGGCCAGGTATTTGTCCAAGTTTCTCCCATGTGATAGCCTGAAAATATGCGCTCGT 378
Db      339 AGAAGGCCAGGTATTTGTCCAAGTTTCTCCCATGTGATAGCCTGAATATGCGCTCAT 398
Qy      379 GGGAAAGGAAAGACTGAC-----CGTCCCGACG 408
Db      399 GGGAAAGGTAAGACTGACTGTCCCGACCGGACATCCCGACCGCGACATCCCGACG 458
Qy      409 CGGACACCCGTAAAGGCTCTGTGCTGAGAGAGATTAAGTAAAGAGAGAAATGCGCTTT 468
Db      459 CGGACACCCGAAAGAGGCTGTGTGCTGAGAGAGATTAAGTAAAGAGAGAAAGG--CTCTTT 515
Qy      469 GCAGTTGAGACAAAGGAAGGACATCTGTCTCTGCTGCTGCCCTGGGCAATGGAATGTC 528
Db      516 GCAGTTGAGTAAAGGAAGGACATCTGTCTCTGCTGCTGCCCTGGGCAATGGAATGTC 575
Qy      529 GGTATAAACCCGATTTGTATGCTCCATCTACTGAGTAGGAAAAACGCTTTAGGACTG 588
Db      576 GGTATAAACCCGATTTGTATGCTTACTTACTGAGTAGGAAAAACATCTTTAGGACTG 635
Qy      589 GAGGTGGGACCTCGCGGGACGAATCTGCTTTGTAAAGCACTGAGATGTTTATGTGTAG 648
Db      636 GAGGTGAACACGCTGCGGGCAATCTGCTTTTATATGCAACGAGATGTTTGTATAGTG 695
Qy      649 CATATCTAAAGACAGACCTTAATCCCTTATACATGTCATGATGCAAGAACTTTGTTTC 708
Db      696 CACATC--AAGGACAGACACTTCTCTTAACTTATTTATATGACACAGAACCTTTGTTTC 752
Qy      709 ACATGTTTGTCTGTACCCCTTCCCACAATTTCTTGTGAACCTTGACACATCCCTTC 768
Db      753 AC--GTTTTCGTGACCCCTTCCCACTATTACCTTATGCGCTGCACATCCCTTC 810
Qy      769 TTGAGAAAACCCACAGATGATCTAGTAATATCTAAGGAACTCAGAGGCTGCG----- 823
Db      811 TCCGAGATGTAGAGATATGATCAATAATCTAGGGGAACTCAGAGACACAGTGTCCCT 870
Qy      824 --GATCTCTCATATGCTGAAGCGTGTGTTCCCGGGTCCCTCTTTCTTTCTTA 877
Db      871 GTAGGTCTCTCGTGTAGCGCGGCTCTCTTGTGGCTCACTTTCTTTCTTATA 926

```

RESULT 8

US-09-598-326-141
Sequence 141, Application US/09596326
Patent No. 6423496
GENERAL INFORMATION:
APPLICANT: Prudakis, Tony N.
Smith, John M.
Reed, Steven G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 247
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed Intellectual Property Law Group PLLC
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/598,326
FILING DATE: 20-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Poter, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 141:
SEQUENCE CHARACTERISTICS:
LENGTH: 9388 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 141:
US-09-598-326-141

Query Match 56.3%; Score 494.8; DB 4; Length 9388;
Best Local Similarity 79.1%; Pred. No. 2,1e-165;
Matches 709; Conservative 0; Mismatches 132; Indels 55; Gaps 8;
Qy 20 GATCAGATTGTTACTGTCGTCTGTGAGAAAGATGATGAGACTCCATTTTGTGA 79
Db 48 GATAGACTGTTACTGTCGTCTGTGAGAAAGATGATGAGACTCCATTTTGTTC 107
Qy 80 TGTACTAAGAAAAATTTCTTCTGCTTGTGATTTCTTATCTATGACCTTACCCCAACC 139
Db 108 TGTACTAAGAAAAATTTCTTCTGCTTGTGATTTCTTATCTATGACCTTACCCCAACC 167
Qy 140 CCGTCTCTCTGAAACATGTCCTGTCTCTCACTCAGGGTTAAATGATTTAAGGGGGTCA 199
Db 168 CTGTGCTCAGAGACATGTCCTGTCTCTCACTCAGGGTTAAATGATTTAAGGGG----- 221
Qy 200 GGATGCTCTTGTAAACAGATGCTTGAAGGACAGATGCTTGAAGGATGATCAACT 259
Db 222 ---TATGCTTTGTTAAAAAGTCTTGAAGTAAATGCTTTGTTAAAGTATCACTCCT 278
Qy 260 CCTAATCTCAATACCCAGGAGACAA-AAAATGCGGAGCGGAGGACCTCTGCT 318
Db 279 CTCTAATCTCAATACCCAGGAGACAAATACATGCGGAGCGGAGGACCTCTGCT 338
Qy 319 AGGAAAGCCAGGTATTTGTCACGTTTCTCCCATGTGATAGCTGAAATATGACCTCTGT 378
Db 339 AGGAAAGCCAGGTATTTGTCACGTTTCTCCCATGTGATAGCTGATATGACCTCTCAT 398
Qy 379 GGGAAAGGAAAGACTAC-----CCTCCCAAGC 408
Db 399 GGGAAAGGAAAGACTACCTGCTCTCCCAAGCCGACATCCCAAGCCGACATCCCAAGC 458
Qy 409 CCGACACCCGTAAGGCTGTGCTGAGAGATTAAGTAAAGAGAAAGATCTCTT 468
Db 459 CCGACACCCGTAAGGCTGTGCTGAGAGATTAAGTAAAGAGAAAG---CCTCTT 515
Qy 469 GCACTGAGACAGAGAGGATCTGCTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 528
Db 516 GCACTGAGAGAGAGAGGATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 575
Qy 529 GGTATAAAGCCGTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 588
Db 576 GGTATAAAGCCGTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 635
Qy 589 GAGTGGGACCTGCGGAGCAATATCTTTGTAAGCACTGAGATTTATGATG 648
Db 636 GAGTGGGACCTGCGGAGCAATATCTTTGTAAGCACTGAGATTTATGATG 695
Qy 649 CATATCTAAGACAGACTTAAATCTTTAATGCTGATGATGACAAAGACTTTGTTTC 708
Db 696 CACATC--AAGGACAGACACT--TTCCCTAACTAATATGACACAGAGACTTTGTTTC 752

Qy 709 ACATGTTGTCGTGACCCCTCTCCCAATATGCTTTGACCTTGACATCCCTC 768
Db 753 AC--GTTTCTGCTGACCCCTCTCCCAATATGCTTTGACCTTGACATCCCTC 810
Qy 769 TTGAGAAACCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 823
Db 811 TCCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 870
Qy 824 --GGATCTCTCAATATGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 877
Db 871 GTAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 926

RESULT 9

US-09-289-198-141

Sequence 141: Application US/09289198

Patent No. 6586570

GENERAL INFORMATION:

APPLICANT: Frudakis, Tony N.

APPLICANT: Smith, John M.

APPLICANT: Reed, Steven G.

APPLICANT: Misher, Lynda

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TITL OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER

FILE REFERENCE: 210121.419C5

CURRENT APPLICATION NUMBER: US/09/289,198

CURRENT FILING DATE: 1999-04-09

EARLIER APPLICATION NUMBER: US 09/062,451

EARLIER FILING DATE: 1998-04-17

EARLIER APPLICATION NUMBER: US 08/991,789

EARLIER FILING DATE: 1997-12-11

EARLIER APPLICATION NUMBER: US 08/838,762

EARLIER FILING DATE: 1997-04-09

EARLIER APPLICATION NUMBER: PCT/US97/00485

EARLIER FILING DATE: 1997-01-10

EARLIER APPLICATION NUMBER: US 08/700,014

EARLIER FILING DATE: 1996-08-20

EARLIER APPLICATION NUMBER: US 08/585,392

EARLIER FILING DATE: 1996-01-01

NUMBER OF SEQ ID NOS: 312

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 141

LENGTH: 9388

TYPE: DNA

ORGANISM: Homo sapien

US-09-289-198-141

Query Match 56.3%; Score 494.8; DB 4; Length 9388;
Best Local Similarity 79.1%; Pred. No. 2,1e-165;
Matches 709; Conservative 0; Mismatches 132; Indels 55; Gaps 8;
Qy 20 GATCAGATTGTTACTGTCGTCTGTGAGAAAGATGATGAGACTCCATTTTGTGA 79
Db 48 GATCAGACTGTTACTGTCGTCTGTGAGAAAGATGATGAGACTCCATTTTGTTC 107
Qy 80 TGTACTAAGAAAAATTTCTTCTGCTTGTGATTTCTTATCTATGACCTTACCCCAACC 139
Db 108 TGTACTAAGAAAAATTTCTTCTGCTTGTGATTTCTTATCTATGACCTTACCCCAACC 167
Qy 140 CCGTCTCTCTGAAACATGTCCTGTCTCTCACTCAGGGTTAAATGATTTAAGGGGGTCA 199
Db 168 CTGTGCTCAGAGACATGTCCTGTCTCTCACTCAGGGTTAAATGATTTAAGGGG----- 221
Qy 200 GGATGCTCTTGTAAACAGATGCTTGAAGGACAGATGCTTGAAGGATGATCAACT 259
Db 222 ---TATGCTTTGTTAAAAAGTCTTGAAGTAAATGCTTTGTTAAAGTATCACTCCT 278
Qy 260 CCTAATCTCAATACCCAGGAGACAA-AAAATGCGGAGCGGAGGACCTCTGCT 318
Db 279 CTCTAATCTCAATACCCAGGAGACAAATACATGCGGAGCGGAGGACCTCTGCT 338
Qy 319 AGGAAAGCCAGGTATTTGTCACGTTTCTCCCATGTGATAGCTGAAATATGACCTCTGT 378

FEATURE:
NAME/KEY: CDS
LOCATION: 119..334
US-09-621-976-2586

Query Match 51.0%; Score 448.6; DB 4; Length 482;
Best Local Similarity 96.9%; Pred. No. 8.3e-150;
Matches 468; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

```
QY 84 CTAAGAAAAATTTCTTCCCTTGAATCTGTATATATGACCTTACCCCAACCCCGT 143
DB 1 CTAAGAAAAATTTCTTCCCTTGAATCTGTATATATGACCTTACCCCAACCCCGT 60
QY 144 GCTCTGAAACATGTGCTGTGCTCACTCAGAGTTAAATGATTAAAGGCGGTGAGAT 203
DB 61 GCTCTGAAACATGTGCTGTGCTCACTCAGAGTTAAATGATTAAAGGCGGTGAGAT 120
QY 204 GTGCTTTGTTAAACAGATGCTTGAAGGAGCATGCTCTTAAGATCACTCCCT 263
DB 121 GTGCTTTGTTAAACAGATGCTTGAAGGAGCATGCTCTTAAGATCACTCCCT 180
QY 264 AATCTCAAGTACCAAGGACACAAAATGCGGAAGGCGGAGGAGCTTGTCTAGGAA 323
DB 181 AATCTCAAGTACCAAGGACACAAAATGCGGAAGGCGGAGGAGCTTGTCTAGGAA 239
QY 324 AGCAGGATTTGTCACAGTTTCTCCCATGTGATAGCTGAATATAGGCTCGTGGAA 383
DB 240 AGCAGGATTTGTCACAGTTTCTCCCATGTGATAGCTGAATATAGGCTCGTGGAA 299
QY 384 GGAAGAACTGACCGTCCCGACCGGACACCGGTAAGGCTGTGCTAGGAGATT 443
DB 300 GGAAGAACTGACCGTCCCGACCGGACACCGGTAAGGCTGTGCTAGGAGATT 359
QY 444 AGTAAAGAGAGAAATGCTCTTTCAGTTGAGACAAAGAGGATCTGTCTCTGC 503
DB 360 AGTAAAGAGAGAAATGCTCTTTCAGTTGAGACAAAGAGGATCTGTCTCTGC 419
QY 504 CTGTCCTGGGCAATGGAATGCTCGGTATTAACCGGATGTATGCTCACTTACTGAG 563
DB 420 CTGTCCTGGGCAATGGAATGCTCGGTATTAACCGGATGTATGCTCACTTACTGAG 479
QY 564 ATA 566
DB 480 ATA 482
```

RESULT 12
US-09-621-976-2859
Sequence 2859, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET 054PR2
CURRENT APPLICATION NUMBER: US/09/621.976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 2859
LENGTH: 488
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 119..334
US-09-621-976-2859

Query Match 50.4%; Score 442.8; DB 4; Length 488;
Best Local Similarity 96.5%; Pred. No. 9.9e-148;
Matches 473; Conservative 1; Mismatches 14; Indels 2; Gaps 2;

```
QY 84 CTAAGAAAAATTTCTTCCCTTGAATCTGTATATATGACCTTACCCCAACCCCGT 143
DB 1 CTAAGAAAAATTTCTTCCCTTGAATCTGTATATATGACCTTACCCCAACCCCGT 60
QY 144 GCTCTGAAACATGTGCTGTGCTCACTCAGAGTTAAATGATTAAAGGCGGTGAGAT 203
DB 61 GCTCTGAAACATGTGCTGTGCTCACTCAGAGTTAAATGATTAAAGGCGGTGAGAT 120
QY 204 GTGCTTTGTTAAACAGATGCTTGAAGGAGCATGCTCTTAAGATCACTCCCT 263
DB 121 GTGCTTTGTTAAACAGATGCTTGAAGGAGCATGCTCTTAAGATCACTCCCT 180
QY 264 AATCTCAAGTACCAAGGACACAAAATGCGGAAGGCGGAGGAGCTTGTCTAGGAA 323
DB 181 AATCTCAAGTACCAAGGACACAAAATGCGGAAGGCGGAGGAGCTTGTCTAGGAA 239
QY 324 AGCAGGATTTGTCACAGTTTCTCCCATGTGATAGCTGAATATAGGCTCGTGGAA 383
DB 240 AGCAGGATTTGTCACAGTTTCTCCCATGTGATAGCTGAATATAGGCTCGTGGAA 299
QY 384 GGAAGAACTGACCGTCCCGACCGGACACCGGTAAGGCTGTGCTAGGAGATT 443
DB 300 GGAAGAACTGACCGTCCCGACCGGACACCGGTAAGGCTGTGCTAGGAGATT 359
QY 444 AGTAAAGAGAGAAATGCTCTTTCAGTTGAGACAAAGAGGATCTGTCTCTGC 503
DB 360 AGTAAAGAGAGAAATGCTCTTTCAGTTGAGACAAAGAGGATCTGTCTCTGC 419
QY 504 CTGTCCTGGGCAATGGAATGCTCGGTATTAACCGGATGTATGCTCACTTACTGAG 563
DB 420 CTGTCCTGGGCAATGGAATGCTCGGTATTAACCGGATGTATGCTCACTTACTGAG 478
QY 564 ATAGGAAAA 573
DB 479 ATAGGAAAA 488
```

RESULT 13
US-09-328-111-600
Sequence 600, Application US/09328111
Patent No. 626233
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Carino, Theodore J.
APPLICANT: Derti, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328.111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 600
LENGTH: 589
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-600

Query Match 41.2%; Score 362; DB 3; Length 589;
Best Local Similarity 78.6%; Pred. No. 8,4e-119;
Matches 469; Conservative 0; Mismatches 120; Indels 8; Gaps 3;

QY 274 ACCGAGGACACAAAACCTGGAGAGCCGACGAGGACCTGCTAGAGAAAGCAGGAT 333
DB 1 ACCGAGGACACAAAACCTGGAGAGCCGACGAGGACCTGCTAGAGAAAGCAGGAT 60
QY 334 TGTCCAAAGTTTCTCCCATGTGATGCTGAAATATGCTGCTGAGAAAGGAAAGACC 333
DB 61 TGTCCAAAGTTTCTCCCATGTGATGCTGAAATATGCTGCTGAGAAAGGAAAGACC 120
QY 394 TGACCGTCCCGACCGGACACCCGTAAGAGGCTGTGCTGAGAGAGATTAGTAAAGAG 453
DB 121 TGACCGTCCCGACCGGACACCCGTAAGAGGCTGTGCTGAGAGAGATTAGTAAAGAG 180
QY 454 GAAGGATGCTCTTGAGATTGAGACAAAGAGAGGATCTGCTCTGCTGCTGCTGCTG 513
DB 181 GAAGG---CCTCTTGAAGTTGAGATGAAGAGAGGATCTGCTCTGCTGCTGCTGCTG 237
QY 514 GCAATGAGATGCTCGGTATTAACCCGATTGTATGCTCATCTACTAGATAGGAGAAA 573
DB 238 GCAATGAGATGCTCGGTATTAACCCGATTGTATGCTCATCTACTAGATAGGAGAAA 297
QY 574 ACCGCTTGAAGGCTGAGAGTGGAGCTGCGGACAGCAATCTGCTTTGTAAGCACTGAG 633
DB 298 ACTGCTTGAAGGCTGAGAGTGGAGCTGCGGACAGCAATCTGCTTTGTAAGCACTGAG 357
QY 634 ATGTTTATGTATGATCATCTAATAAGCAGCACTTAATCCTTAATGATGTCTATGATG 693
DB 358 ATGTTTATGTATGATCA---AAAAGCAGCAAGCCCTTTCTTACCTGCTTATATGATG 415
QY 694 CAAAGACCTTGTTCACATGTTTGTCTGAGACCTCTCCGACAAATGTCTGTGAGACC 753
DB 416 CAAAGACCTTGTTCACATGTTTGTCTGAGACCTCTCCGACAAATGTCTGTGAGACC 475
QY 754 TGACACATCCCTCTCTGAGAAACCCACAGATGATCAATAATACTAAGGAGACTCA 813
DB 476 CCACA---TCTCTTTTGAAGAGTGAATATGATCAATAATACTAAGGAGACTNAG 532
QY 814 GAGGCTGGCGGATCTTCATATGCTGAACGCTGCTCCCGGCTCCCTCTCTTCT 870
DB 533 ANAGTGGCGCGTGAAGTCTTAATATCTGAACGCGAGTCTCCGCTGCTGCTGCTGCTGCT 589

RESULT 14

US-08-991-789A-7
Sequence 7, Application US/08991789A
Patent No. 6225054

GENERAL INFORMATION:

APPLICANT: Frudakis, Tony N.
Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed IP Law Group

STREET: 701 Fifth Avenue, Suite 6300

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/991,789A

FILING DATE: 11-Dec-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1086 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear

US-08-991-789A-7
SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Query Match 40.5%; Score 355.6; DB 3; Length 1086;
Best Local Similarity 80.4%; Pred. No. 2.5e-116;
Matches 532; Conservative 0; Mismatches 81; Indels 49; Gaps 8;

QY 20 GATCAGATTGTTACTGTGCTGTGTGTAAGAAAGTAGACATGAGAGATTCATTGTTA 79
DB 54 GATCAGACTGTACTGTGCTATGTAGAAAGATGACATAGAGATTCATTGTTGTT 113
QY 80 TGTACTAAGAAAATTCTTCTGCTTGAATTTCTTTATCTATGACCTTACCCCAACC 139
DB 114 TGTACTAAGAAAATTCTTCTGCTTGAATTTCTTTATCTGTAACCTTACCCCAACC 173
QY 140 CCGTGTCTCTGTAACATGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 199
DB 174 CTGTGCTCAGAGACATGT 227
QY 200 GATGTGCTTTGTTAAACAGATGCTTGAAGGACATGCTCTTAAAGTCAATCACT 259
DB 228 ---TATGCTTTGTTAAAGTGTCTGTAAGATATATGCTTGTAAAGTCAATCACT 284
QY 260 CCTATCTCAATGATCCAGGACACAA---AACTGGGAGAGGCGGAGGAGGAGGAGGAGG 318
DB 285 CTCTAATCTCAATGATCCAGGACACAAATGATGCTGGAAGGCGGAGGAGGAGGAGGAG 344
QY 319 AGGAAAGCCAGGATGCTCAACGTTTCTCCCATGTGATGCTGATGCTGATGCTGATGCTG 378
DB 345 AGGAAAGCCAGGATGCTCAACGTTTCTCCCATGTGATGCTGATGCTGATGCTGATGCTG 404
QY 379 GGGAGGAGAAAGACCTGACCG-----TCCCCAGC 408
DB 405 GGGAGGAGTAAAGCTGACGCTGCTCCGACGACATCCCGACCGACATCCCGACG 464
QY 409 CCGACACCCGTAAGAGGCTGTGCTGAGAGGATTTAGTAAAGAGAGGAGGAGGAGGAGG 468
DB 465 CCGACACCCGTAAGAGGCTGTGCTGAGAGGATTTAGTAAAGAGAGGAGGAGGAGGAGG 519
QY 469 GCAGTTGAGACAGAGAGAGGATGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528
DB 520 TGCATTGAGTAAAGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 578
QY 529 GGTATTAACCCGATTGTATGCTCATCTAGAG-ATAGGAGAAACCGCTTGAAGGCT 587
DB 579 GGTATTAACCCGATTGTATGCTCATCTAGAGATAGAGAGAAACATCTTGAAGGCT 638
QY 588 GAGAGTGGAGCTGCGGAGAGCAATCTGCTTTGTAAGAGAGAGATGTTTATGCTAT 647
DB 639 GAGAGTGGAGCAACCTGCGGAG-ATAGTGTCTTTATGAC- GAGATGTTTATTAATT 696
QY 648 GC 649
DB 697 GC 698

RESULT 15

US-09-062-451-7
Sequence 7, Application US/09062451
Patent No. 6344550

GENERAL INFORMATION:

this page blank (us010)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 21:22:12 ; Search time 77.9883 Seconds
(without alignments)
8170.833 Million cell updates/sec

Title: US-10-016-604-155

Perfect score: 150
Sequence: 1 gagataggggaaaaacgcctc.....atgcaagaccttgcac 150

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	100.0	150	6	ABX04696 Human end
2	150	100.0	150	9	ADCI15773 Putative
3	150	100.0	319	6	ABX04570 Human end
4	150	100.0	319	6	ADCI16256 Human her
5	150	100.0	408	6	ABX04571 Human end
6	150	100.0	408	6	ADCI16257 Human her
7	150	100.0	879	6	ABX04572 Human end
8	150	100.0	879	6	ADCI16258 Human her
9	150	100.0	968	6	ABX04691 Human end
10	150	100.0	968	6	ADCI15768 Human her
11	148.4	98.9	2431	4	AAK94873 Human ful
12	148.4	98.9	6556	4	AAK70301 Human imm
13	148.4	98.9	35414	3	AAAD00147 TR12 reia
14	148.4	98.9	222930	6	ABK84349 Human CDN
15	146.8	97.9	968	6	ABX04695 Human end
16	146.8	97.9	968	6	ABX04694 Human end
17	146.8	97.9	968	6	ABX04693 Human end
18	146.8	97.9	968	9	ADCI15772 HML-2 LTR
19	146.8	97.9	968	9	ADCI15770 HML-2 LTR
20	146.8	97.9	968	9	ADCI15771 HML-2 LTR
21	146.8	97.9	32249	4	AAI05336 Human rep
22	146.8	97.9	32249	4	ABL98205 Human tes
23	146	97.3	2525	3	AAZ98026 Human sec

C	24	146	97.3	2525	4	AAI11639 Human sec
C	25	146	97.3	2525	6	ABK69735 Human sec
C	26	145.2	96.8	1659	5	AA66662 DNA encod
C	27	145.2	96.8	1701	5	AA83303 DNA encod
C	28	145.2	96.8	1704	5	AA83549 DNA encod
C	29	145.2	96.8	1782	5	AA879381 DNA encod
C	30	145.2	96.8	1826	5	AA870436 DNA encod
C	31	145.2	96.8	3658	5	AA870654 DNA encod
C	32	145.2	96.8	9179	6	ABX04612 Human end
C	33	145.2	96.8	9659	6	AAI57703 Human pro
C	34	145.2	96.8	9659	8	ADA02672 Human PRL
C	35	145.2	96.8	9659	9	ADB72410 Human PRL
C	36	145.2	96.8	154902	6	ABQ88198 Human ost
C	37	143.6	95.7	150	9	ADCI15738 TAR nucle
C	38	143.6	95.7	962	6	ABX04692 Human end
C	39	143.6	95.7	962	9	ADCI15769 HML-2 LTR
C	40	143.6	95.7	139904	6	ABK83562 Human .CDN
C	41	143.6	95.7	144792	9	ADC87620 Human GPC
C	42	143.6	95.7	349981	9	ADC87619 Human GPC
C	43	140.4	93.6	975	7	ACC43227 Nucleotid
C	44	140.4	93.6	2689	7	ACC43223 Nucleotid
C	45	140.4	93.6	6020	7	ACC43236 Nucleotid

ALIGNMENTS

RESULT 1.

ABX04696 standard; DNA, 150 BP.

AC ABX04696;

DT 14-JAN-2003 (first entry)

XX Human endogenous retrovirus k (herv-k) 5' RUS region.

XX Human; endogenous retrovirus; herv; prostate cancer; testicular cancer; multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;

KW cancer; transgenic animal; ds.

OS Human endogenous retrovirus.

PN WO200246477-A2.

XX 07-DEC-2001; 2001WO-US047824.

PF 07-DEC-2000; 2000US-0251830P.

PR 07-DEC-2001; 2001US-00016604.

PA (CHIR) CHIRON CORP.

PI Garcia P, Hardy SF, Williams LT, Escobedo J;

DR WPI; 2002-691475/74.

XX Novel isolated polypeptides useful for diagnosis of prostate cancer.

PS Claim 4; Page 149; 152pp; English.

XX The invention describes novel isolated polypeptides (I, Ib) useful for
XX diagnosing prostate cancer comprising obtaining a patient sample
XX containing prostate cells and detecting the presence or absence of an
XX expression product of a HML-2 endogenous retrovirus in a patient sample.
XX Polynucleotides associated with (I) are useful for diagnosis or treatment
XX of testicular cancer, multiple sclerosis or insulin-dependent diabetes
XX mellitus. An inhibitor of a HML-2 protease and a transdominant negative
XX mutant of HML-2 GORP are also useful in the manufacture of a medicament
XX for treating prostate cancer. (I) and (Ib) are useful for generating
XX antibodies specific to the polypeptides associated with cancer, as
XX targets for therapeutic intervention, and in immunising a transgenic

CC animal. This sequence represents a region of human endogenous retrovirus
CC (herv) of the HML-2 sub-group
XX
SQ Sequence 150 BP; 43 A; 29 C; 37 G; 41 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 6; Length 150;
Best Local Similarity 100.0%; Pred. No. 4,1e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGGAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGGCAGCAATACCTGCTT 60
DB 1 GAGATAGGGAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGGCAGCAATACCTGCTT 60
QY 61 GTAAGCAGCTGAGATGTTTATGTATGATGATCTAAAGCAGACCTTAATCCTTTAC 120
DB 61 GTAAGCAGCTGAGATGTTTATGTATGATGATCTAAAGCAGACCTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAAAGACCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAAAGACCTTTGTTTAC 150

RESULT 2

ADCI5773
ID ADCI5773 standard; DNA; 150 BP.

AC ADCI5773;

DT 18-DEC-2003 (first entry)

DE Putative TAR of herv-k(hml-2.hom) nucleotide sequence SEQ ID NO:49.

XX prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
KM immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
KW multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.

XX Homo sapiens.

OS WO2003050258-A2.

PN 19-JUN-2003.

PD 09-DEC-2002; 2002WO-US039344.

PF 07-DEC-2001; 2001US-00016604.

PR 07-DEC-2001; 2001US-0340664P.

PR 07-DEC-2001; 2001WO-US047824.

PR 12-JUN-2002; 2002US-0388046P.

PA (CHIR) CHIRON CORP.

PI Garcia P, Hardy SF, Williams LT, Escobedo J;

DR WPI; 2003-577296/54.

XX Diagnosing prostate cancer, comprises detecting the presence or absence
PT of HML-2 (human endogenous retrovirus) expression product in a patient
PT sample.

PS Disclosure; SEQ ID NO 49; 117bp, English.

XX The present invention describes a method for diagnosing prostate cancer,
CC comprising detecting the presence or absence of HML-2 expression product
CC in a patient sample, where the expression product is produced by a
CC splicing event in which the 5' region and start codon of the env coding
CC region are joined to a downstream coding region in the reading frame +2
CC relative to that of env. Also described: (1) an isolated polynucleotide;
CC (2) an isolated polypeptide; (3) an antibody that binds to the
CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
CC sequences have cytosolic, neuroprotective, antidiabetic and
CC immunostimulant activities, and can be used in vaccines and in gene
CC therapy. A polynucleotide, polypeptide or antibody of the invention can
CC be used in the manufacture of a medicament for preventing or treating

CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
CC dependent diabetes mellitus. The present sequence represents a sequence
CC which is used in the exemplification of the present invention.

XX
SQ Sequence 150 BP; 43 A; 29 C; 37 G; 41 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 9; Length 150;
Best Local Similarity 100.0%; Pred. No. 4,1e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGGAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGGCAGCAATACCTGCTT 60
DB 1 GAGATAGGGAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGGCAGCAATACCTGCTT 60
QY 61 GTAAGCAGCTGAGATGTTTATGTATGATGATCTAAAGCAGACCTTAATCCTTTAC 120
DB 61 GTAAGCAGCTGAGATGTTTATGTATGATGATCTAAAGCAGACCTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAAAGACCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAAAGACCTTTGTTTAC 150

RESULT 3

ABX04570
ID ABX04570 standard; DNA; 319 BP.

AC ABX04570;

DT 14-JAN-2003 (first entry)

DE Human endogenous retrovirus k (herv-k) R region.

XX Human, endogenous retrovirus; herv; prostate cancer; testicular cancer;
KM multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
KW cancer; transgenic animal; ds.

XX Human endogenous retrovirus.

OS WO200246477-A2.

PN 13-JUN-2002.

PD 07-DEC-2001; 2001WO-US047824.

PF 07-DEC-2000; 2000US-0251830P.

PR 07-DEC-2001; 2001US-00016604.

PA (CHIR) CHIRON CORP.

PI Garcia P, Hardy SF, Williams LT, Escobedo J;

DR WPI; 2002-691475/74.

XX Novel isolated polypeptides useful for diagnosis of prostate cancer.

PS Disclosure; Page 129; 152bp; English.

XX The invention describes novel isolated polypeptides (I, II) useful for
CC diagnosing prostate cancer comprising obtaining a patient sample
CC containing prostate cells and detecting the presence or absence of an
CC expression product of a HML-2 endogenous retrovirus in a patient sample.
CC Polynucleotides associated with (I) are useful for diagnosis or treatment
CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes
CC mellitus. An inhibitor of a HML-2 protease and a transdominant negative
CC mutant of HML-2 CORF are also useful in the manufacture of a medicament
CC for treating prostate cancer. (I) and (II) are useful for generating
CC antibodies specific to the polypeptides associated with cancer, as
CC targets for therapeutic intervention, and in immunising a transgenic
CC animal. This sequence represents a region of human endogenous retrovirus
CC (herv) of the HML-2 sub-group
XX
SQ Sequence 319 BP; 79 A; 83 C; 69 G; 88 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 6; Length 319;
Best Local Similarity 100.0%; Pred. No. 5.4e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGGGAAAAACCCCTTTAGGCTGAGAGTGGACCTTGGCGGACGCAATATCTGCTTT 60
DB 1 GAGATAGGGGAAAAACCCCTTTAGGCTGAGAGTGGACCTTGGCGGACGCAATATCTGCTTT 60
QY 61 GTAAGACACTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 61 GTAAGACACTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150

RESULT 4
ADCl6256
ID ADCl6256 standard; DNA; 319 BP.
XX ADCl6256;
XX
XX 18-DEC-2003 (first entry)
XX
XX
DE Human herv-k(hml-2.hom) R region nucleotide sequence SEQ ID NO:3.
XX
XX
KW prostate cancer; HML-2; env; cytostatic; neuroprotective; antidiabetic;
KW immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
KW multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.
XX
XX Homo sapiens.
XX
XX WO2003050258-A2.
XX
XX 19-JUN-2003.
XX
XX
XX 09-DEC-2002; 2002WO-US039344.
XX
XX 07-DEC-2001; 2001US-00016604.
XX 07-DEC-2001; 2001US-034064P.
XX 07-DEC-2001; 2001WO-US047824.
XX 12-JUN-2002; 2002US-0388046P.
XX
XX (CHIR) CHIRON CORP.
XX
XX Garcia P, Hardy SF, Williams LT, Escobedo J;
XX
XX WPI; 2003-577296/54.
XX
XX
PT Diagnosing prostate cancer, comprises detecting the presence or absence
PT of HML-2 (human endogenous retrovirus) expression product in a patient
PT sample.
XX
XX Disclosure; SEQ ID NO 3; 117pp; English.

CC The present invention describes a method for diagnosing prostate cancer,
CC comprising detecting the presence or absence of HML-2 expression product
CC in a patient sample, where the expression product is produced by a
CC splicing event in which the 5' region and start codon of the env coding
CC region are joined to a downstream coding region in the reading frame +2
CC relative to that of env. Also described: (1) an isolated polynucleotide;
CC (2) an isolated polypeptide; (3) an antibody that binds to the
CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
CC sequences have cytostatic, neuroprotective, antidiabetic and
CC immunostimulant activities, and can be used in vaccines and in gene
CC therapy. A polynucleotide, polypeptide or antibody of the invention can
CC be used in the manufacture of a medicament for preventing or treating
CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
CC dependent diabetes mellitus. The present sequence represents a sequence
CC which is used in the exemplification of the present invention.

SQ Sequence 319 BP; 79 A; 83 C; 69 G; 88 T; 0 U; 0 Other.

Query Match 100.0%; Score 150; DB 9; Length 319;
Best Local Similarity 100.0%; Pred. No. 5.4e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGGGAAAAACCCCTTTAGGCTGAGAGTGGACCTTGGCGGACGCAATATCTGCTTT 60
DB 1 GAGATAGGGGAAAAACCCCTTTAGGCTGAGAGTGGACCTTGGCGGACGCAATATCTGCTTT 60
QY 61 GTAAGACACTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 61 GTAAGACACTGAGATGTTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150

RESULT 5
ABX04571
ID ABX04571 standard; DNA; 408 BP.
XX ABX04571;
XX
XX 14-JAN-2003 (first entry)
XX
XX
DE Human endogenous retrovirus k (herv-k) RUS region.
XX
XX
KW Human; endogenous retrovirus; herv; prostate cancer; testicular cancer;
KW multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
KW cancer; transgenic animal; ds.
XX
XX
XX Human endogenous retrovirus.
XX
XX WO200246477-A2.
XX
XX 13-JUN-2002.
XX
XX
XX 07-DEC-2001; 2001WO-US047824.
XX
XX 07-DEC-2000; 2000US-0251830P.
XX 07-DEC-2001; 2001US-00016604.
XX
XX (CHIR) CHIRON CORP.
XX
XX Garcia P, Hardy SF, Williams LT, Escobedo J;
XX
XX WPI; 2002-691475/74.
XX
XX
PT Novel isolated polypeptides useful for diagnosis of prostate cancer.
PT Disclosure; Page 129; 152pp; English.

CC The invention describes novel isolated polypeptides (I, Ib) useful for
CC diagnosing prostate cancer comprising obtaining a patient sample
CC containing prostate cells and detecting the presence or absence of an
CC expression product of a HML-2 endogenous retrovirus in a patient sample.
CC Polynucleotides associated with (I) are useful for diagnosis or treatment
CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes
CC mellitus. An inhibitor of a HML-2 protease and a transdominant negative
CC mutant of HML-2 CORP are also useful in the manufacture of a medicament
CC for treating prostate cancer. (I) and (Ib) are useful for generating
CC antibodies specific to the polypeptides associated with cancer, as
CC targets for therapeutic intervention, and in immunising a transgenic
CC animal. This sequence represents a region of human endogenous retrovirus
CC (herv) of the HML-2 sub-group
XX
XX SQ Sequence 408 BP; 97 A; 113 C; 83 G; 115 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 6; Length 408;
Best Local Similarity 100.0%; Pred. No. 5.9e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGGGCAATACTGCTT 60
DB 1 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGGGCAATACTGCTT 60
QY 61 GTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTTAC 120
DB 61 GTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGAACCCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAGAACCCTTTGTTTAC 150

RESULT 6

ADCl6257
ID ADCl6257 standard; DNA; 408 BP.

AC ADCl6257;

DT 18-DEC-2003 (first entry)

XX Human herV-k(hml-2.hom) RUS region nucleotide sequence SEQ ID NO:4.

XX prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
XX immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
XX multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.

OS Homo sapiens.

PN WO2003050258-A2.

PD 19-JUN-2003.

PF 09-DEC-2002; 2002WO-US039344.

PR 07-DEC-2001; 2001US-00016604.

PR 07-DEC-2001; 2001US-034064P.

PR 07-DEC-2001; 2001WO-US047824.

PR 12-JUN-2002; 2002US-0388046P.

PA (CHIR) CHIRON CORP.

PI Garcia P, Hardy SF, Williams LT, Escobedo J;

XX WPI; 2003-577296/54.

PT Diagnosing prostate cancer, comprises detecting the presence or absence
PT of HML-2 (human endogenous retrovirus) expression product in a patient
PT sample.

PS Disclosure; SEQ ID NO 4; 117bp; English.

XX The present invention describes a method for diagnosing prostate cancer,
XX comprising detecting the presence or absence of HML-2 expression product
XX in a patient sample, where the expression product is produced by a
XX splicing event in which the 5' region and start codon of the env coding
XX region are joined to a downstream coding region in the reading frame +2
XX relative to that of env. Also described: (1) an isolated polynucleotide;
XX (2) an isolated polypeptide; (3) an antibody that binds to the
XX polypeptide; and (4) an immunogenic composition. Prostate cancer related
XX sequences have cytosolic, neuroprotective, antidiabetic and
XX immunostimulant activities, and can be used in vaccines and in gene
XX therapy. A polynucleotide, polypeptide or antibody of the invention can
XX be used in the manufacture of a medicament for preventing or treating
XX prostate, breast or testicular cancer, multiple sclerosis or insulin-
XX dependent diabetes mellitus. The present sequence represents a sequence
XX which is used in the exemplification of the present invention.

XX Sequence 408 BP; 97 A; 113 C; 83 G; 115 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 9; Length 408;
Best Local Similarity 100.0%; Pred. No. 5.9e-41;

Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGGGCAATACTGCTT 60
DB 1 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGGGCAATACTGCTT 60
QY 61 GTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTTAC 120
DB 61 GTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTTAC 120
QY 121 ATTGCTATGATGCAAGAACCCTTTGTTTAC 150
DB 121 ATTGCTATGATGCAAGAACCCTTTGTTTAC 150

RESULT 7

ABX04572
ID ABX04572 standard; DNA; 879 BP.

AC ABX04572;

DT 14-JAN-2003 (first entry)

XX Human endogenous retrovirus k (herV-k) U3R region.

XX Human; endogenous retrovirus; herV; prostate cancer; testicular cancer;
XX multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
XX cancer; transgenic animal; ds.

OS Human endogenous retrovirus.

PN WO200246477-A2.

PD 13-JUN-2002.

PF 07-DEC-2001; 2001WO-US047824.

PR 07-DEC-2000; 2000US-0251830P.

PR 07-DEC-2001; 2001US-00016604.

PA (CHIR) CHIRON CORP.

PI Garcia P, Hardy SF, Williams LT, Escobedo J;

XX WPI; 2002-691475/74.

PT Novel isolated polypeptides useful for diagnosis of prostate cancer.

PS Claim 4; Page 129; 152bp; English.

XX The invention describes novel isolated polypeptides (I, Ib) useful for
XX diagnosing prostate cancer comprising obtaining a patient sample
XX containing prostate cancer cells and detecting the presence or absence of an
XX expression product of a HML-2 endogenous retrovirus in a patient sample.
XX Polynucleotides associated with (I) are useful for diagnosis or treatment
XX of testicular cancer, multiple sclerosis or insulin-dependent diabetes
XX mellitus. An inhibitor of a HML-2 protease and a transdominant negative
XX mutant of HML-2 cORF are also useful in the manufacture of a medicament
XX for treating prostate cancer. (I) and (Ib) are useful for generating
XX antibodies specific to the polypeptides associated with cancer, as
XX targets for therapeutic intervention, and in immunising a transgenic
XX animal. This sequence represents a region of human endogenous retrovirus
XX (herV) of the HML-2 sub-group

XX Sequence 879 BP; 231 A; 212 C; 211 G; 225 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 6; Length 879;
Best Local Similarity 100.0%; Pred. No. 7.9e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGGGCAATACTGCTT 60
DB 561 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGGGCAATACTGCTT 620

Db 61 GTAAAGCACTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
Qy |||||
Db 621 GTAAAGCACTGAGATGTTATGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 680
Qy 121 ATTGCTATGATGCAAGAACCTTTGTTTAC 150
|||
Db 681 ATTGCTATGATGCAAGAACCTTTGTTTAC 710
|||

RESULT 8
ADCl6258
ID ADCl6258 standard; DNA; 879 BP.
XX
AC ADCl6258;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human herV-K(hml-2.hom) U3R region nucleotide sequence SEQ ID NO:5.
XX
KM prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
KM immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
KM multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003050258-A2.
XX
PD 19-JUN-2003.
XX
PF 09-DEC-2002; 2002WO-US039344.
XX
PR 07-DEC-2001; 2001US-00016604.
PR 07-DEC-2001; 2001US-034064P.
PR 07-DEC-2001; 2001WO-US047824.
PR 12-JUN-2002; 2002US-0388046P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX
DR WPI; 2003-577296/54.
XX
PT Diagnosing prostate cancer, comprises detecting the presence or absence
PT of HML-2 (human endogenous retrovirus) expression product in a patient
PT sample.
XX
PS Disclosure; SEQ ID NO 5; 117bp; English.
XX
CC The present invention describes a method for diagnosing prostate cancer,
CC comprising detecting the presence or absence of HML-2 expression product
CC in a patient sample, where the expression product is produced by a
CC splicing event in which the 5' region and start codon of the env coding
CC region are joined to a downstream coding region in the reading frame +2
CC relative to that of env. Also described: (1) an isolated polynucleotide;
CC (2) an isolated polypeptide; (3) an antibody that binds to the
CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
CC sequences have cytosolic, neuroprotective, antidiabetic and
CC immunostimulant activities, and can be used in vaccines and in gene
CC therapy. A polynucleotide, polypeptide or antibody of the invention can
CC be used in the manufacture of a medicament for preventing or treating
CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
CC dependent diabetes mellitus. The present sequence represents a sequence
CC which is used in the exemplification of the present invention.
XX
SQ Sequence 879 BP; 231 A; 212 C; 211 G; 225 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 9; Length 879;
Best Local Similarity 100.0%; Pred. No. 7.9e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAAAGCCCTTAGGGCTGAGGTGGAGACTGCGGAGCAATACCTGCTT 60
|||

Db 561 GAGATAGGAAAAAAGCCCTTAGGGCTGAGGTGGAGACTGCGGAGCAATACCTGCTT 620
Qy 61 GTAAAGCACTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
|||
Db 621 GTAAAGCACTGAGATGTTATGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 680
Qy 121 ATTGCTATGATGCAAGAACCTTTGTTTAC 150
|||
Db 681 ATTGCTATGATGCAAGAACCTTTGTTTAC 710
|||

RESULT 9
ABX04691
ID ABX04691 standard; DNA; 968 BP.
XX
AC ABX04691;
XX
DT 14-JAN-2003 (first entry)
XX
DE Human endogenous retrovirus k (herV-K) long terminal repeat #1.
XX
KM Human endogenous retrovirus; herV; prostate cancer; testicular cancer;
KM multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
KM cancer; transgenic animal; ds.
XX
OS Human endogenous retrovirus.
XX
PN WO200246477-A2.
XX
PD 13-JUN-2002.
XX
PF 07-DEC-2001; 2001WO-US047824.
XX
PR 07-DEC-2000; 2000US-0251830P.
PR 07-DEC-2001; 2001US-00016604.
XX
PA (CHIR) CHIRON CORP.
XX
PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX
DR WPI; 2002-691475/74.
XX
PT Novel isolated polypeptides useful for diagnosis of prostate cancer.
XX
PS Claim 18; Page 148; 152pp; English.
XX
CC The invention describes novel isolated polypeptides (I, Ib) useful for
CC diagnosing prostate cancer comprising obtaining a patient sample
CC containing prostate cells and detecting the presence or absence of an
CC expression product of a HML-2 endogenous retrovirus in a patient sample.
CC Polynucleotides associated with (I) are useful for diagnosis or treatment
CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes
CC mellitus. An inhibitor of a HML-2 protease and a transdominant negative
CC mutant of HML-2 CORF are also useful in the manufacture of a medicament
CC for treating prostate cancer. (I) and (Ib) are useful for generating
CC antibodies specific to the polypeptides associated with cancer, as
CC targets for therapeutic intervention, and in immunising a transgenic
CC animal. This sequence represents a region of human endogenous retrovirus
CC (herV) of the HML-2 sub-group
XX
SQ Sequence 968 BP; 249 A; 242 C; 225 G; 252 T; 0 U; 0 Other;

Query Match 100.0%; Score 150; DB 6; Length 968;
Best Local Similarity 100.0%; Pred. No. 8.2e-41;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAAAGCCCTTAGGGCTGAGGTGGAGACTGCGGAGCAATACCTGCTT 60
|||
Db 561 GAGATAGGAAAAAAGCCCTTAGGGCTGAGGTGGAGACTGCGGAGCAATACCTGCTT 620
|||
Qy 61 GTAAAGCACTGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
|||
Db 621 GTAAAGCACTGAGATGTTATGTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 680
|||

Db 1670 ATTGCTATGATGCAGAACCTTGTTCAC 1699

RESULT 12
ID AAK70301/C
ID AAK70301 standard; DNA; 6556 BP.

XX AAK70301;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25113.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236378P.
PR 29-SEP-2000; 2000US-0236380P.
PR 29-SEP-2000; 2000US-0236389P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249256P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.

XX 14-AUG-2002 (first entry)
DT Human cDNA differentially expressed in granulocytic cells #920.
XX
XX Human; 89; granulocytic cell; DNA chip; bacterial infection;
KM viral infection; parasitic infection; protozoal infection;
KM fungal infection; sterile inflammatory disease; psoriasis;
KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KM cardiac reperfusion injury; renal reperfusion injury; AIDS;
KM adult respiratory distress syndrome; inflammatory bowel disease;
KM Crohn's disease; ulcerative colitis; periodontal disease;
KM granulocyte activation; chronic inflammation; allergy.
XX Homo sapiens.
PN WO200228999-A2.
PD 11-APR-2002.
XX 03-OCT-2001; 2001WO-US030821.
PF 03-OCT-2000; 2000US-0237189P.
XX (GENE-) GENE LOGIC INC.
PA Beazer-Barclay Y, Weltsman SM, Yamaga S, Vockley J;
PI WPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
PS Claim 1; SEQ ID NO 920; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, AIDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 222930 BP; 53286 A; 56778 G; 57221 G; 55643 T; 0 U; 2 Other;

Query Match 98.9%; Score 148.4; DB 6; Length 222930;
Best Local Similarity 99.3%; Pred. No. 2,2e-39;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GAGATAGGAAAAACCGCTTAGGCTGAGGTGGACCTGCGGACGAACTACTGCTTT 60
Db 98271 GAGATAGGAAAAACCGCTTAGGCTGAGGTGGACCTGCGGACGAACTACTGCTTT 98212
Qy 61 GTAAAGCATGAGATGTTTATGTATGATCATATTTAAAGCAGCATTAATCTTTAC 120
Db 98211 GTAAAGCATGAGATGTTTATGTATGATCATATTTAAAGCAGCATTAATCTTTAC 98152
Qy 121 ATTGCTATGATGCAAGACCTTTGTCAC 150
Db 98151 ATTGCTATGATGCAAGACCTTTGTCAC 98122
RESULT 15
ABX04695
ID ABX04695 standard; DNA; 968 BP.
XX
XX ABX04695;
AC
XX
XX 14-JAN-2003 (first entry)
DT
XX
XX Human endogenous retrovirus k (herv-k) long terminal repeat #5.
XX
XX Human; endogenous retrovirus; herv; prostate cancer; testicular cancer;
KM multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
KM cancer; transgenic animal; de.
XX
XX Human endogenous retrovirus.
OS
XX
XX WO200246477-A2.
PN
XX
XX 13-JUN-2002.
PD
XX
XX 07-DEC-2001; 2001WO-US047824.
PF
XX
XX 07-DEC-2000; 2000US-0251830P.
PR
XX
XX 07-DEC-2001; 2001US-00016604.
PA (CHIR) CHIRON CORP.
XX
XX Garcia P, Hardy SF, Williams LT, Escobedo J;
PI WPI; 2002-691475/74.
DR
XX
XX Novel isolated polypeptides useful for diagnosis of prostate cancer.
PT
XX
XX Claim 18; Page 149; 152pp; English.
PS
XX
XX The invention describes novel isolated polypeptides (I, Ib) useful for
CC diagnosing prostate cancer comprising obtaining a patient sample
CC containing prostate cells and detecting the presence or absence of an
CC expression product of a HML-2 endogenous retrovirus in a patient sample.
CC Polynucleotides associated with (I) are useful for diagnosis or treatment
CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes
CC mellitus. An inhibitor of a HML-2 protease and a transdominant negative
CC mutant of HML-2 CORP are also useful in the manufacture of a medicament
CC for treating prostate cancer. (I) and (Ib) are useful for generating
CC antibodies specific to the polypeptides associated with cancer, as
CC targets for therapeutic intervention, and in immunising a transgenic
CC animal. This sequence represents a region of human endogenous retrovirus
CC (herv) of the HML-2 sub-group
XX
XX Sequence 968 BP; 250 A; 237 C; 225 G; 256 T; 0 U; 0 Other;

Query Match 97.9%; Score 146.8; DB 6; Length 968;
Best Local Similarity 98.7%; Pred. No. 1e-39;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY      1 GAGATAGGGAAAAACCGCCTTTAGGGCTGAGGTGGACCTGCGGGCAGCAATCTGCTT 60
      |||||||
Db      561 GAGATAGGGAAAAACCGCCTTTAGGGCTGAGGTGGACCTGCGGGCAGCAATCTGCTT 620
      |||||||
QY      61  GTAAGCACTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTAC 120
      |||||||
Db      621 GTAAGCATTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTAC 680
      |||||||
QY      121 ATTGCTATGATGCAAAGACCTTGTTCAC 150
      |||||||
Db      681 CTGTCTATGATGCAAAGACCTTGTTCAC 710
      |||||||
```

Search completed: February 27, 2004, 22:44:04
Job time : 79.9883 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 27, 2004, 22:29:18 ; Search time 609.038 Seconds
(without alignments)
7354.759 Million cell updates/sec

Title: US-10-016-604-155

Sequence: 1 gagatagggaacacgcctc.....atgcaagaccttgcac 150

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_earth.*
2: em_earth.*
3: em_earth.*
4: em_earth.*
5: em_earth.*
6: em_earth.*
7: em_earth.*
8: em_earth.*
9: gb_earth.*
10: gb_earth.*
11: gb_earth.*
12: gb_earth.*
13: gb_earth.*
14: gb_earth.*
15: gb_earth.*
16: em_earth.*
17: em_earth.*
18: em_earth.*
19: em_earth.*
20: em_earth.*
21: em_earth.*
22: em_earth.*
23: em_earth.*
24: em_earth.*
25: em_earth.*
26: em_earth.*
27: em_earth.*
28: gb_earth.*
29: gb_earth.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	148.4	98.9	444	9	A1066622 o282b02.x
C 2	148.4	98.9	1204	12	BM907033 AGENCOURT
C 3	147.4	98.3	460	14	H80802 Y664d11.r1
C 4	146.8	97.9	356	10	BF737628 PM1-KT003

Result No.	Score	Query Match	Length	ID	Description
C 5	146.8	97.9	448	9	AA393030 z69f06.x
C 6	146.8	97.9	488	12	B1048071 PM2-ST030
C 7	146.8	97.9	524	12	BM693416 UI-E-CK1
C 8	146.8	97.9	576	13	B0632008 UI-H-FE1
C 9	146.8	97.9	633	13	BM668693 UI-E-CK1
C 10	146.8	97.9	689	13	B0575726 UI-H-E21
C 11	146.8	97.9	706	13	B0618928 UI-H-FH1
C 12	146.8	97.9	707	12	BM981862 UI-CF-EN1
C 13	145.2	96.8	438	10	BF918605 CM0-NT013
C 14	145.2	96.8	704	12	BM989989 UI-H-D10
C 15	145.2	96.8	818	13	B0216036 AGENCOURT
C 16	143.6	95.7	452	9	AA573159 m51e12.s
C 17	143.6	95.7	1698	11	BC036118 Homo sapi
C 18	142	94.7	874	14	CD558661 AGENCOURT
C 19	140.4	93.6	372	13	B0360083 CM0-HN020
C 20	140.4	93.6	586	9	AA708837 z69f06.x
C 21	140.4	93.6	689	14	CD684466 EST1006.h
C 22	140.4	93.6	798	12	B1758163 B03023895
C 23	138.8	92.5	409	13	BQ312183 QV3-BN037
C 24	137.2	91.5	649	9	AL049055 DKFZP434P
C 25	136.4	90.4	541	14	W52520 zc54b08.r1
C 26	135.6	90.4	351	12	B1038269 RC5-NT026
C 27	135.6	90.4	471	9	AA703380 zj12b06.s
C 28	135.6	90.4	570	12	BQ027101 UI-H-ED0
C 29	135.6	90.4	670	12	BQ006641 UI-H-E11
C 30	135.6	90.4	897	13	BX374012 BX374012
C 31	135.6	90.4	954	13	BX327468 BX327468
C 32	135.2	90.1	232	10	BF948093 M33-NN022
C 33	135.2	90.1	322	14	T05809 EST03698 Fe
C 34	133.8	89.2	292	9	AA076809 7B03C10 C
C 35	133.6	89.1	975	9	AL565978 AL565978
C 36	133.2	88.8	485	12	BM667549 UI-E-DX0
C 37	133.2	88.8	678	14	CA449778 UI-H-E11
C 38	131.8	87.9	305	13	BQ318376 CM4-CT051
C 39	131.8	87.9	329	13	BQ318375 CM4-CT051
C 40	131.8	87.9	670	14	CB850840 UI-CF-EN1
C 41	131.4	87.6	300	9	AA716377 z69f06.x
C 42	130.4	86.9	339	9	AA633703 ag87b07.s
C 43	130.2	86.8	622	12	BM985460 UI-CF-EC1
C 44	130	86.7	632	10	AW963304 EST375377
C 45	127.6	85.1	376	13	BQ318373 CM4-CT051

ALIGNMENTS

RESULT 1
LOCUS A1066622/c
DEFINITION o282b02.x1 Soares senescent fibroblasts NBHSF Homo sapiens CDNA
clone IMAGE:1681803 3' similar to contains LTR5.t1 LTR5 repetitive
element ;, mRNA sequence.

ACCESSION A1066622
VERSION A1066622
KEYWORDS A1066622.1 GI:3367324
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 444)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapdb-remail.nih.gov
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 615
Seq Error: 0.00
Seq primer: -40m13 fwd: RT from Amersham.
Location/Qualifiers
1. .444
/organism="Homo sapiens"

FEATURES

source

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1681803"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares senescent fibroblasts_NHSP"
/note="Vector: p773D (Pharmacia) with a modified
polylinker V_type: phagemid; Site 1: Not I; Site 2: Eco
RI; 1st strand CDNA was primed with a Not I - oligo (dT)
primer [5',
TGTTACCAATCTGAAGTGGAGGGCGGCAATTTTCTTTTCTTTT 3']
double-stranded CDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773D vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 98.9%; Score 148.4; DB 9; Length 444;
Best Local Similarity 99.3%; Pred. No. 1.5e-35;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGATAGGGAAAAACCCCTTAGGGCTGAGGTGGACCTCGGGGCAATACTGCTT 60
DB 323 GAGATAGGGAAAAACCCCTTAGGGCTGAGGTGGACCTCGGGGCAATACTGCTT 264
QY 61 GTAAAGCACTGAGATGTTTATGTATGATCTAATAAGACAGCACTTAATCCTTTAC 120
DB 263 GTAAAGCACTGAGATGTTTATGTATGATCTAATAAGACAGCACTTAATCCTTTAC 204
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 203 ATTGCTATGATGCAAGACCTTTGTTTAC 174

RESULT 2

BM907033 1204 bp mRNA linear EST 12-MAR-2002
LOCUS AGENCOURT 6622286 NIH_MGC_124 Homo sapiens CDNA clone IMAGE:5735089
DEFINITION 5', mRNA sequence.

ACCESSION BM907033 GI:19357412
VERSION BM907033.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 1204)
NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov

Tissue Procurement: Invitrogen

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov
Plate: LLM12741 row: h column: 02

High quality sequence start: 3
High quality sequence stop: 598.

FEATURES

source

1..1204
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5735089"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_124"

/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: EcoRV
(destroyed); Site 2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."

ORIGIN

Query Match 98.9%; Score 148.4; DB 12; Length 1204;
Best Local Similarity 99.3%; Pred. No. 2.1e-35;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGATAGGGAAAAACCCCTTAGGGCTGAGGTGGACCTCGGGGCAATACTGCTT 60
DB 383 GAGATAGGGAAAAACCCCTTAGGGCTGAGGTGGACCTCGGGGCAATACTGCTT 442
QY 61 GTAAAGCACTGAGATGTTTATGTATGATCTAATAAGACAGCACTTAATCCTTTAC 120
DB 443 GTAAAGCACTGAGATGTTTATGTATGATCTAATAAGACAGCACTTAATCCTTTAC 502
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 503 ATTGCTATGATGCAAGACCTTTGTTTAC 532

RESULT 3

H80802/c 460 bp mRNA linear EST 09-NOV-1995
LOCUS YB64011.r1 Soares retina N2B4HR Homo sapiens CDNA clone
DEFINITION IMAGE:219573 5' similar to contains LTRs repetitive element ;, mRNA
sequence.

ACCESSION H80802 GI:1058891
VERSION H80802.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 460)

REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Matra, M.,
Patterson, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

The Mashu-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: eglew@wustl.edu

High quality sequence stops: 325

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

Insert length: 903 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 325.
Location/Qualifiers
1..460
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3847566"
/db_xref="taxon:9606"
/clone="IMAGE:219573"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares retina N2B4HR"

Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5', TGTACCAATCTGAAGTGGAGCGCCGCCCAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to C05, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 97.9%; Score 146.8; DB 9; Length 448;
Best Local Similarity 98.7%; Pred. No. 4.6e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAACCCCTTAGGCTGAGTGGACCTGCGGACGCAATCTGCTT 60
Db 209 GAGATAGGAAAAACCCCTTAGGCTGAGTGGACCTGCGGACGCAATCTGCTT 268
Qy 61 GTAAGACCTGAGATGTTATGTGTATGCAATCTAAAGACAGCACTTAATCCTTAC 120
Db 269 GTAAGACCTGAGATGTTATGTGTATGCAATCTAAAGACAGCACTTAATCCTTAC 328
Qy 121 ATTGCTATGATGCAAGACCTTGTTCAC 150
Db 329 ATTGCTATGATGCAAGACCTTGTTCAC 358

RESULT 6
Bi048071/c 488 bp mRNA linear EST 14-JUN-2001
LOCUS Bi048071
DEFINITION PM2-ST0303-230301-011-h03 ST0303 Homo sapiens cDNA, mRNA sequence.
ACCESSION Bi048071
VERSION Bi048071.1 GI:14454693
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,M., Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baita,S.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,D.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202863
PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&ct=PM2-ST0303-230301-011-h03&ct3=2001-03-23&ct4=1)
Seq primer: puc 18 forward
High quality sequence stop: 7
High quality sequence stop: 488.

Location/Qualifiers
1. .488
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="adult"
/clone_id="ST0303"
/note="Organ: stomach; Vector: puc18; site_1: Sma1;

FEATURES

source

Site 2: Sma1: A mini-library was made by cloning products derived from ONESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 97.9%; Score 146.8; DB 12; Length 488;
Best Local Similarity 98.7%; Pred. No. 4.8e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAACCCCTTAGGCTGAGTGGACCTGCGGACGCAATCTGCTT 60
Db 232 GAGATAGGAAAAACCCCTTAGGCTGAGTGGACCTGCGGACGCAATCTGCTT 173
Qy 61 GTAAGACCTGAGATGTTATGTGTATGCAATCTAAAGACAGCACTTAATCCTTAC 120
Db 172 GTAAGACCTGAGATGTTATGTGTATGCAATCTAAAGACAGCACTTAATCCTTAC 113
Qy 121 ATTGCTATGATGCAAGACCTTGTTCAC 150
Db 112 ATTGCTATGATGCAAGACCTTGTTCAC 83

RESULT 7
BM693416 524 bp mRNA linear EST 28-FEB-2002
LOCUS BM693416
DEFINITION UI-E-CK1-afm-b-10-0-UI_r1 UI-E-CK1 Homo sapiens cDNA clone
ACCESSION UI-E-CK1-afm-b-10-0-UI 5', mRNA sequence.
VERSION BM693416.1 GI:19006674
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548

COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.regen.com).

The following repetitive elements were found in this cDNA sequence: 1-521, >LTR5#LTR/Retroviral
Seq primer: M13 Reverse.
Location/Qualifiers
1. .524
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afm-b-10-0-UI"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="PH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-CK1"
/note="Organ: eye; Vector: pT73-Pac (pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CK1 is a normalized cDNA library containing the

FEATURES
source

1. .524
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afm-b-10-0-UI"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="PH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-CK1"
/note="Organ: eye; Vector: pT73-Pac (pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-CK1 is a normalized cDNA library containing the

following clone(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 97.9%; Score 146.8; DB 12; Length 524;
Best Local Similarity 98.7%; Pred. No. 4.9e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGATAGGAGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGCGAGCAATACCTGCTT 60
DB 301 GAGATAGGAGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGCGAGCAATACCTGCTT 360
QY 61 GTAAGCACTGAGATGTTTATGTGTATGATATCTAAAGCAGACCTTAATCCTTTAC 120
DB 361 GTAAGCACTGAGATGTTTATGTGTATGATATCTAAAGCAGACCTTAATCCTTTAC 420
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 421 ATTGCTATGATGCAAGACCTTTGTTTAC 450

RESULT 8

BU632008/c 576 bp mRNA linear EST 23-SEP-2002
LOCUS UI-H-FEI-bea-c-09-0-UI.s1 NCI CGAP FEI Homo sapiens CDNA clone
DEFINITION UI-H-FEI-bea-c-09-0-UI 3', mRNA sequence.
ACCESSION BU632008
VERSION BU632008.1 GI:23299263
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: sgabbs-remail.nih.gov
Tissue Procurement: James Martin
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this CDNA
sequence: 18-576, >LTR5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

source 1.576
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FEI-bea-c-09-0-UI"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP FEI"
/notes="Organ: Chondrosarcoma; Vector: p773-Pac
(Pharmacia) with a modified polylinker; Site_1: EcoR I;

Site 2: Not I; NCI CGAP FEI is a normalized CDNA library derived from a pool of mRNA obtained from 3 cell lines from grade II chondrosarcoma tissues. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGCTACGAGC. The cell lines were provided by Dr James Martin from the University of Iowa.
TAG_TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG_LIB=UI-H-FEI
TAG_SEQ=CGCTACGAGC"

ORIGIN

Query Match 97.9%; Score 146.8; DB 13; Length 576;
Best Local Similarity 98.7%; Pred. No. 5.1e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAGATAGGAGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGCGAGCAATACCTGCTT 60
DB 334 GAGATAGGAGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGCGAGCAATACCTGCTT 275
QY 61 GTAAGCACTGAGATGTTTATGTGTATGATATCTAAAGCAGACCTTAATCCTTTAC 120
DB 274 GTAAGCACTGAGATGTTTATGTGTATGATATCTAAAGCAGACCTTAATCCTTTAC 215
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 214 ATTGCTATGATGCAAGACCTTTGTTTAC 185

RESULT 9

BM668693/c 633 bp mRNA linear EST 27-FEB-2002
LOCUS UI-E-CK1-afm-b-10-0-UI.s2 UI-E-CK1 Homo sapiens CDNA clone
DEFINITION UI-E-CK1-afm-b-10-0-UI 3', mRNA sequence.
ACCESSION BM668693
VERSION BM668693.1 GI:18978590
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Normalization and subfraction: two approaches to facilitate gene
JOURNAL discovery
MEDLINE Genome Res. 6 (9), 791-806 (1996)
PUBMED 97044477
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this CDNA
sequence: 18-633, >LTR5#LTR/Retroviral (matched complement)
Seq primer: M13 Forward
POLYA=yes.

FEATURES

location/Qualifiers

```

source
1. .633
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afm-b-10-0-UI"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_id="UI-E-CK1"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: Ecor I; Site 2: Not I;
UI-E-CK1 is a normalized cDNA library containing the
following tissue(s): Retina Foveal and Macular. The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an Ecor I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GTCC. This library was created for the program,
Gene Discovery in the Visual System, supported by National
Eye Institute (NEI).
TAG_TISSUE=Foveal and Macular Retina
TAG_LIB=UI-E-CK1
TAG_SEQ=GTCC"

```

ORIGIN

Query Match	97.9%	Score 146.8	DB 12	Length 633
Best Local Similarity	98.7%	Pred. No. 5.2e-35		
Matches 148	Conservative 0	Mismatches 2	Indels 0	Gaps 0

QY	1	GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGACCTGCGGACAGCAATACTGCTTT	60
Db	334	GAGATATGGGAAAAACCGCCTTAGGGCTGGAGGTGGACCTGCGGACAGCAATACTGCTTT	275
QY	61	GTTAAGCACTGAGATGTTTATGTGTATGATATCTTAAAGACAGCACTTAATCTTTAC	120
Db	274	GTTAAGCACTGAGATGTTTATGTGTATGATATCTTAAAGACAGCACTTAATCTTTAC	215
QY	121	ATTGCTATGATGCAAGACCTTTTTCAC	150
Db	214	ATTGCTATGATGCAAGACCTTTTTCAC	185

RESULT 10	LOCUS	DEFINITION
B0575726/c	B0575726	689 bp mRNA linear EST 19-JUN-2002
	UI-H-E21-bbg-f-06-0-UI.g1	NCI CGAP Ch2 Homo sapiens CDNA clone
	UI-H-E21-bbg-f-06-0-UI 3'	mRNA sequence.

ACCESSION	B0575726.1	GI:21479043
VERSION		
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE 1 (bases 1 to 689)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

The following repetitive elements were found in this cDNA
 Sequence: 16-689, >LTRs#LTR/Retroviral (matched complement
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES	Location/Qualifiers
source	1. .689

```

SOURCE
1. .689
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="U1-H-E21-bbg-f-06-0-U1"
/tissue_type="Chondrosarcoma Grade II"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Ch2"
/notes="Organ: Left Pelvis; Vector: pP7T3-Pac (Pharmacia)
with a modified polylinker; Site_1: SCOR I; Site_2: Not I
NCI_CGAP_Ch2 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma Grade II. The library
was constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pP7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
TAGTCACGCT."
TAG_TISSUE="grade-2-chondrosarcoma"
TAG_LIB="U1-H-E21"
TAG_SEQ="ATCTAATAG"

```

ORIGIN

Query Match	97.9%	Score 146.8;	DB 13;	Length 689;
Best Local Similarity	98.7%	Pred. No. 5.4e-35;		
Matches 148; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

QY 1 GAGATGCGAAAAAACCCGCTTAGGGCTGAGAGGTGGACCTGC GGAGCAGCAATACTGTTT 60
|||
|||
Db 271 GGAATGGGAAAAAACCGCCCTTAGGGCTGAGAGGTGGACTTCGCGGGCAGCAATCTGCTTT 212
|||
|||
QY 61 GPAAGACCTGAGTCTATGTGTGATGCAATCTAAAGCACGCACTTAATCCTTTAC 120
|||
|||
Db 211 GPAAGACAATGAGATTATGTGTGATGCAATCTCAAAGCACGCACTTAATCCTTTAC 152
|||
|||
QY 121 ATTGCTATGATGACAAAGACCTTTGTTTAC 150
|||
|||
Db 151 ATTGCTATGATGCCAAGACCTTTGTTTAC 122
|||
|||

RESULT 11					
BUG18928/c					
LOCUS	BUG18928	706 bp	mRNA	linear	EST_23-SEP-2002
DEFINITION	BU-H-FH1-bf1-b-11-0-UI.s1 NCT_CGAP_FH1 Homo sapiens CDNA clone UI-H-FH1-bf1-b-11-0-UI 3, mRNA sequence.				

ACCESSION	Bu615928	
VERSION	Bu615928.1	GI:23285143
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE	1 (phases 1 to 706)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D.

Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu. The following repetitive elements were found in this cDNA sequence: 18-706, >LTR5#LTR/Retroviral (matched complement). Seq primer: M13 FORWARD

FEATURES

source

Location/Qualifiers
1..706
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-R-FH1-bf1-b-11-0-UI"
/tissue_type="Cell Line"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP FH1"
/note="Organ: Chondrosarcoma; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP FH1 is a normalized cDNA library obtained from a cell line derived from grade I chondrosarcoma tissue. The library was constructed and normalized according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGATCCGGC. The cell line was provided by Dr. James Martin from the University of Iowa.
TAG_TISSUE=Human Chondrosarcoma Cell Line C98 - Grade 1
Chondrosarcoma
TAG_LIB=UI-R-FH1
TAG_SEQ=AGATCCGGC"

ORIGIN

Query Match 97.9%; Score 146.8; DB 13; Length 706;
Best Local Similarity 98.7%; Pred. No. 5.5e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GAGTATGAGGAAAACCGCTTAGGCTGAGAGTGAGGACCTGCGGACGACATACCTGCTT 60
DB 272 GAGTATGAGGAAAACCGCTTAGGCTGAGAGTGAGGACCTGCGGACGACATACCTGCTT 213
QY 61 GTAAGCACTGAGATGTTTATGTATGATCATATCTAAAGCAGACACTTAATCCTTTAC 120
DB 212 GTAAGCACTGAGATGTTTATGTATGATCATATCTAAAGCAGACACTTAATCCTTTAC 153
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 152 ATTGCTATGATGCAAGACCTTTGTTTAC 123

RESULT 12
BM981862/c 707 bp mRNA linear EST 21-FEB-2003
LOCUS
DEFINITION
UI-CF-EN1-adj-g-12-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
BM981862
ACCESSION
BM981862.1 GI:19604781
VERSION
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 707)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovey
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477

PUBMED

COMMENT

8889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.research.com) or from Open Biosystems (www.openbiosystems.com).
The following repetitive elements were found in this cDNA sequence: 17-707, >LTR5#LTR/Retroviral (matched complement). Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

source

Location/Qualifiers
1..707
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-adj-g-12-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.
TAG_TISSUE=Human Lung Epithelial Cell lines untreated LPS
6hr to LPS 24h
TAG_LIB=UI-CF-EN1
TAG_SEQ=CTGCTCAGGT"

ORIGIN

Query Match 97.9%; Score 146.8; DB 12; Length 707;
Best Local Similarity 98.7%; Pred. No. 5.5e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GAGTATGAGGAAAACCGCTTAGGCTGAGAGTGAGGACCTGCGGACGACATACCTGCTT 60
DB 298 GAGTATGAGGAAAACCGCTTAGGCTGAGAGTGAGGACCTGCGGACGACATACCTGCTT 239
QY 61 GTAAGCACTGAGATGTTTATGTATGATCATATCTAAAGCAGACACTTAATCCTTTAC 120
DB 238 GTAAGCACTGAGATGTTTATGTATGATCATATCTAAAGCAGACACTTAATCCTTTAC 179
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 178 ATTGCTATGATGCAAGACCTTTGTTTAC 149

RESULT 13
BF918605 438 bp mRNA linear EST 19-JAN-2001
LOCUS
DEFINITION
CMO-NT0133-131000-614-b04 NT0133 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BF918605.1 GI:12314493
VERSION

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 438)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcritpome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PubMed 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CMO&ct2=CMO-NT0133-131000-614-B04&ct3=2000-10-13&ct4=1)
Seq primer: puc 18 forward
High quality sequence start: 3
High quality sequence stop: 438.
Location/Qualifiers
1. 438
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NT0133"
/note="Organ: nervous tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
ORIGIN
Query Match 96.8%; Score 145.2; DB 10; Length 438;
Best Local Similarity 98.0%; Pred. No. 1.4e-34;
Matches 147; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAGATTAGGAAAAACCGCTTAGGGCTGAGGTGGGACCTGCGGGAGCAATCTGCTTT 60
DB 267 GAGATTAGGAAAAACCGCTTAGGGCTGAGGTGGGACCTGCGGGAGCAATCTGCTTT 326
QY 61 GTAAGACATGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 327 GTAAGACATGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 386
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 387 ATTGCTATGATGCAAGACCTTTGTTTAC 416

RESULT 14
BM989989 704 bp mRNA linear EST 17-JUN-2002
LOCUS UI-H-DIO-ato-h-05-0-UI.s1.NCI.CGAP_D10 Homo sapiens cDNA clone
DEFINITION IMAGE:5862196 3', mRNA sequence.
ACCESSION BM989989
VERSION BM989989.1 GI:19709378
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 704)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@remail.nih.gov
Tissue Procurement: Dr. Jose Mercende
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LIN at: http://image.llnl.gov
The following repetitive elements were found in this cDNA sequence: 16-704, >LIR5#LTR/Retroviral (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1. 704
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5862196"
/cvsue_type="Lung Focal Fibrosis"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP D10"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP D10 is a cDNA library containing the following tissue(s): A pool of Lung Focal Fibrosis. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA, synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ATACGGCGTC.
TAG TISSUE=Lung with fibrosis
TAG LIB=UI-H-DIO
TAG_SEQ=ATACGGCGTC"

FEATURES
source
1. 704
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5862196"
/cvsue_type="Lung Focal Fibrosis"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP D10"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP D10 is a cDNA library containing the following tissue(s): A pool of Lung Focal Fibrosis. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA, synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ATACGGCGTC.
TAG TISSUE=Lung with fibrosis
TAG LIB=UI-H-DIO
TAG_SEQ=ATACGGCGTC"

ORIGIN
Query Match 96.8%; Score 145.2; DB 12; Length 704;
Best Local Similarity 98.0%; Pred. No. 1.7e-34;
Matches 147; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAGATTAGGAAAAACCGCTTAGGGCTGAGGTGGGACCTGCGGGAGCAATCTGCTTT 60
DB 270 GAGATTAGGAAAAACCGCTTAGGGCTGAGGTGGGACCTGCGGGAGCAATCTGCTTT 211
QY 61 GTAAGACATGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 120
DB 210 GTAAGACATGAGATGTTATGTATGATCATATCTAAAGACAGCACTTAATCCTTTAC 151
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 150 CTTGCTATGATGCAAGACCTTTGTTTAC 121

RESULT 15
BQ216036 818 bp mRNA linear EST 02-MAY-2002
LOCUS BQ216036
DEFINITION AGENCOURT_7553694 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6046113
ACCESSION BQ216036
KEYWORDS EST.

VERSION BQ216036.1 GI:20397447
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Straubeberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM13291 row: 9 column: 10
 High quality sequence stop: 577.

FEATURES

source

Location/Qualifiers

1..818

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6046113"

/issue_type="embryonal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_id="NIH_MGC_92"

/note="Organ: testis; Vector: PCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 2.5 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match

96.8%; Score 145.2; DB 13; Length 818;

Best Local Similarity 98.0%; Pred. No. 1.8e-34; Mismatches 3; Indels 0; Gaps 0;

Matches 147; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGACGACATATCTGCTT 60
 |||||
 Db 156 GAGATAGGAAAAACCGCTTAGGGCTGAGGTGGACCTGCGGACGACATATCTGCTT 97
 |||||
 QY 61 GTAAGACATGAGATGTTTATGTATGATGATATCTAAAGACACGACTTAATCTTTAC 120
 |||||
 Db 96 GTAAGACATGAAATGTTTATGTATGATGATATCTAAAGACACGACTTAATCTTTAC 37
 |||||
 QY 121 ATTGTATGATGCAAGACCTTGTTCAC 150
 |||||
 Db 36 CTGTCTATGATGCAAGACCTTGTTCAC 7
 |||||

Search completed: February 28, 2004, 01:07:17
 Job time : 613.038 secs

This Page Blank (USPO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 22:34:13 ; Search time 18.2216 Seconds
(without alignments)
4568.356 Million cell updates/sec

Title: US-10-016-604-155

Perfect score: 1 gagataggggaaaaacgcct.....atgcaagaccttcttccac 150

Sequence: 1 gagataggggaaaaacgcct.....atgcaagaccttcttccac 150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Issued_Patents_MA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCUTS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	146	97.3	2525	US-09-482-273-20	Sequence 20, Appl
2	115.6	77.1	246240	US-08-724-394A-20	Sequence 20, Appl
3	115.6	77.1	246240	US-08-724-394A-21	Sequence 21, Appl
4	115.6	77.1	246240	US-08-724-394A-22	Sequence 22, Appl
5	106.8	71.2	176373	US-09-128-155-17	Sequence 17, Appl
6	92.8	61.9	589	US-09-328-111-600	Sequence 600, App
7	76	50.7	9388	US-08-991-789A-141	Sequence 141, App
8	76	50.7	9388	US-09-062-451-141	Sequence 141, App
9	76	50.7	9388	US-09-598-326-141	Sequence 141, App
10	76	50.7	9388	US-09-289-198-141	Sequence 141, App
11	76	50.7	9388	US-09-429-755-141	Sequence 141, App
12	51.6	34.4	2622	US-09-904-615-45	Sequence 45, Appl
13	39.2	26.1	519	US-08-991-789A-240	Sequence 240, App
14	39.2	26.1	519	US-09-062-451-240	Sequence 240, App
15	39.2	26.1	519	US-09-598-326-240	Sequence 240, App
16	39.2	26.1	519	US-09-289-198-240	Sequence 240, App
17	39.2	26.1	519	US-09-429-755-240	Sequence 240, App
18	38.4	25.6	90541	US-09-759-359A-3	Sequence 3, Appl
19	33.8	22.5	1086	US-08-991-789A-7	Sequence 7, Appl
20	33.8	22.5	1086	US-09-062-451-7	Sequence 7, Appl
21	33.8	22.5	1086	US-09-598-326-7	Sequence 7, Appl
22	33.8	22.5	1086	US-09-289-198-7	Sequence 7, Appl
23	33.8	22.5	1086	US-09-429-755-7	Sequence 7, Appl
24	31.2	20.8	459	US-09-621-976-2273	Sequence 2273, App
25	29.8	19.9	6157	US-08-956-171E-184	Sequence 184, Appl
26	28.4	18.9	7163	US-08-961-527-67	Sequence 67, Appl
27	27.8	18.5	615	US-09-328-111-32	Sequence 32, Appl

C 28	27.6	18.4	417	4	US-09-280-116-69	Sequence 69, Appl
29	27.6	18.4	2963	3	US-09-232-200-60	Sequence 60, Appl
30	27.6	18.4	2963	4	US-09-232-197-60	Sequence 60, Appl
31	27.6	18.4	2963	4	US-09-232-201-60	Sequence 60, Appl
32	27.6	18.4	2963	4	US-09-232-195-60	Sequence 60, Appl
33	27.4	18.3	2241	4	US-09-134-001C-377	Sequence 377, App
34	27.4	18.3	2411	3	US-09-188-930-75	Sequence 75, Appl
35	27.4	18.3	2411	3	US-09-188-930-756	Sequence 256, App
36	27.4	18.3	2411	4	US-09-312-283C-75	Sequence 256, App
37	27.4	18.3	2411	4	US-09-312-283C-256	Sequence 52, Appl
38	27.2	18.1	1837	2	US-08-933-750C-52	Sequence 52, Appl
39	27.2	18.1	1837	3	US-09-234-613-52	Sequence 10674, A
40	27	18.0	531	4	US-09-621-976-10674	Sequence 1332, Ap
C 41	26.8	17.9	591	4	US-09-702-705-1332	Sequence 1336, Ap
C 42	26.8	17.9	591	4	US-09-702-705-1536	Sequence 1332, Ap
C 43	26.8	17.9	591	4	US-09-736-457-1332	Sequence 1336, Ap
C 44	26.8	17.9	591	4	US-09-736-457-1536	Sequence 1332, Ap
C 45	26.8	17.9	591	4	US-09-614-124B-1332	Sequence 1332, Ap

ALIGNMENTS

RESULT 1
US-09-482-273-20/c
; Sequence 20, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482, 273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092, 921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092, 922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092, 956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 2525
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1354)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-482-273-20

Query Match 97.3%; Score 146; DB 4; Length 2525;
Best Local Similarity 95.3%; Pred. No. 2.8e-41;
Matches 143; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGATAGGAAAAACCCCTTAGGGCTGAGTGGAGCTGCGGAGCATATCTGCTT 60
DB 1210 GAGATAGGAAAAACCCCTTAGGGCTGAGTGGAGCTGCGGAGCATATCTGCTT 1151
QY 61 GTAAGACATGAGATGTTTATGTATGATGATCTTAAAGACAGCACTTATCTTTAC 120
DB 1150 GTAAGACATGAGATGTTTATGTATGATGATCTTAAAGACAGCACTTATCTTTAC 1091

QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
Db 1090 MTGTGTATGATGCAAGACCTTTGTTTCAC 1061

RESULT 2
US-08-724-394A-20

; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Laufer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; TITLE OF INVENTION: Sequences and Antibodies Thereo
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Flets, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-20

Query Match 77.1%; Score 115.6; DB 2; Length 246240;
Best Local Similarity 90.0%; Pred. No. 9.9e-30;

Matches 135; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1 GAGATAGGAAAAACCGCTTAGGCGTGAAGTGGGACCTGCGGGCAGCAATACCTGCTT 60
Db 105006 GAGATAGGAAAAACCTGCTTAGGCGTGAAGTGGGACCTGCGGGCAGCAATACCTGCTT 105065
QY 61 GTAAGCACTGAGATGTTTATGTATGATGATATCTAAAGCAGCACTTAATCTTTAC 120
Db 105066 TCAAGTCATGAGATGTTTATGTATGATGATATCT-AAAGCAGCACTTAATCTTTAC 105124
QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
Db 105125 CTGTGTATGATGCAAGACCTTTGTTTCAC 105154

RESULT 3
US-08-724-394A-21

; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Laufer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; TITLE OF INVENTION: Sequences and Antibodies Thereo
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Flets, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-21

Query Match 77.1%; Score 115.6; DB 2; Length 246240;
Best Local Similarity 90.0%; Pred. No. 9.9e-30;

Matches 135; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1 GAGATAGGAAAAACCGCTTAGGCGTGAAGTGGGACCTGCGGGCAGCAATACCTGCTT 60
Db 105006 GAGATAGGAAAAACCTGCTTAGGCGTGAAGTGGGACCTGCGGGCAGCAATACCTGCTT 105065
QY 61 GTAAGCACTGAGATGTTTATGTATGATGATATCTAAAGCAGCACTTAATCTTTAC 120
Db 105066 TCAAGTCATGAGATGTTTATGTATGATGATATCT-AAAGCAGCACTTAATCTTTAC 105124
QY 121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
Db 105125 CTGTGTATGATGCAAGACCTTTGTTTCAC 105154

RESULT 4
US-08-724-394A-22

; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:

APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Laufer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchinashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Filtes, Renee A. 35,136
REGISTRATION NUMBER: 017957-000100
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H-CONTIG"
US-08-724-394A-22
Query Match 77.1%; Score 115.6; DB 2; Length 246240;
Best Local Similarity 90.0%; Pred. No. 9.9e-30;
Matches 135; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
QY 1 GAGATTAGGAAAAACCGCTTAGGGCTGAGAGTGGGACCTTGGCGGACGACATATCTGCTT 60
DB 105006 GAGATTAGGAAAAACCGCTTAGGGCTGAGAGTGGGACCTTGGCGGACGACATATCTGCTT 105065
QY 61 GTAAGACCTGAGATGTTTATGTATGATGATCTATAAAGACAGACACTTAATCCTTTAC 120
DB 105066 TCAAGCTATTGAGATGTTTATGTATGATGATCTT-AAAAGACAGACACTTAATCCTTTAC 105124
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 105125 CTGCTTATGATGACAGACCTTTGTTTAC 105154
RESULT 5
US-09-128-155-17
Sequence 17, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Tang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155

CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 176373
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(176373)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17
Query Match 71.2%; Score 106.8; DB 3; Length 176373;
Best Local Similarity 90.7%; Pred. No. 1e-26;
Matches 136; Conservative 0; Mismatches 12; Indels 2; Gaps 2;
QY 1 GAGATTAGGAAAAACCGCTTAGGGCTGAGAGTGGGACCTTGGCGGACGACATATCTGCTT 60
DB 175231 GAGATTAGGAAAAACCGCTTAGGGCTGAGAGTGGGACCTTGGCGGACGACATATCTGCTT 175289
QY 61 GTAAGACCTGAGATGTTTATGTATGATGATCTATAAAGACAGACACTTAATCCTTTAC 120
DB 175290 TTAAGACCTGAGATGTTTATGTATGATGATCTT-AAAAGACAGACACTTAATCCTTTAC 175348
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 175349 CTGCTTATGATGACAGACCTTTGTTTAC 175378
RESULT 6
US-09-328-111-600
Sequence 600, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 600
LENGTH: 589
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(589)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-600
Query Match 61.9%; Score 92.8; DB 3; Length 589;
Best Local Similarity 80.7%; Pred. No. 6.1e-23;
Matches 121; Conservative 0; Mismatches 27; Indels 2; Gaps 1;

STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/598,326
FILING DATE: 20-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Poter, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 141:
SEQUENCE CHARACTERISTICS:
LENGTH: 938 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 141:
US-09-598-326-141

Query Match 50.7%; Score 76; DB 4; Length 938;
Best Local Similarity 78.0%; Pred. No. 1.6e-16;
Matches 117; Conservative 0; Mismatches 30; Indels 3; Gaps 2;
DB 1 GAGATAGGAGAAAAACCCCTTAGGCTGAGGTGGAGACTCGCGGACGACATACTGCTTT 60
DB 608 GAGATAGGAGAAAAACCTTAGGCTGAGGTGGAGAGACGCTGCGGACATACTGCTCT 667
QY 61 GTAAGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCTTTAC 120
DB 668 TTAATGACCGAGATGTTTATGATGACATC--AAGCAGACGACCT-TTCCTTAAA 724
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 725 CTTATTATGACAGACCTTTGTTTAC 754

RESULT 10
US-09-289-198-141
Sequence 141, Application US/09289198
Patent No. 6586570
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
APPLICANT: Misher, Lynda
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.419C5
CURRENT APPLICATION NUMBER: US/09/289,198
CURRENT FILING DATE: 1999-04-09
EARLIER APPLICATION NUMBER: US 09/062,451
EARLIER FILING DATE: 1998-04-17
EARLIER APPLICATION NUMBER: US 08/991,789
EARLIER FILING DATE: 1997-12-11
EARLIER APPLICATION NUMBER: US 08/838,762
EARLIER FILING DATE: 1997-04-09
EARLIER APPLICATION NUMBER: PCT/US97/00485
EARLIER FILING DATE: 1997-01-10
EARLIER APPLICATION NUMBER: US 08/700,014
EARLIER FILING DATE: 1996-08-20
EARLIER APPLICATION NUMBER: US 08/585,392
EARLIER FILING DATE: 1996-01-01

NUMBER OF SEQ ID NOS: 312
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 141
LENGTH: 938
TYPE: DNA
ORGANISM: Homo sapien
US-09-289-198-141

Query Match 50.7%; Score 76; DB 4; Length 938;
Best Local Similarity 78.0%; Pred. No. 1.6e-16;
Matches 117; Conservative 0; Mismatches 30; Indels 3; Gaps 2;
DB 1 GAGATAGGAGAAAAACCCCTTAGGCTGAGGTGGAGACTCGCGGACGACATACTGCTTT 60
DB 608 GAGATAGGAGAAAAACCTTAGGCTGAGGTGGAGAGACGCTGCGGACATACTGCTCT 667
QY 61 GTAAGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCTTTAC 120
DB 668 TTAATGACCGAGATGTTTATGATGACATC--AAGCAGACGACCT-TTCCTTAAA 724
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 725 CTTATTATGACAGACCTTTGTTTAC 754

RESULT 11
US-09-429-755-141
Sequence 141, Application US/09429755A
Patent No. 6656480
GENERAL INFORMATION:
APPLICANT: Fridakis, Tony N.
APPLICANT: Smith, John M.
APPLICANT: Reed, Steven G.
APPLICANT: Misher, Lynda
APPLICANT: Retter, Marc W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.419C6
CURRENT APPLICATION NUMBER: US/09/429,755A
CURRENT FILING DATE: 1999-10-28
NUMBER OF SEQ ID NOS: 315
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 141
LENGTH: 938
TYPE: DNA
ORGANISM: Homo sapien
US-09-429-755-141

Query Match 50.7%; Score 76; DB 4; Length 938;
Best Local Similarity 78.0%; Pred. No. 1.6e-16;
Matches 117; Conservative 0; Mismatches 30; Indels 3; Gaps 2;
DB 1 GAGATAGGAGAAAAACCCCTTAGGCTGAGGTGGAGACTCGCGGACGACATACTGCTTT 60
DB 608 GAGATAGGAGAAAAACCTTAGGCTGAGGTGGAGAGACGCTGCGGACATACTGCTCT 667
QY 61 GTAAGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCTTTAC 120
DB 668 TTAATGACCGAGATGTTTATGATGACATC--AAGCAGACGACCT-TTCCTTAAA 724
QY 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
DB 725 CTTATTATGACAGACCTTTGTTTAC 754

RESULT 12
US-09-904-615-45/C
Sequence 45, Application US/09904615
Patent No. 6566325
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 49 Human Secreted Proteins

FILE REFERENCE: P2032P1
CURRENT APPLICATION NUMBER: US/09/904,615
CURRENT FILING DATE: 2001-07-16
PRIOR APPLICATION NUMBER: 09/511,554
PRIOR FILING DATE: 2000-02-23
PRIOR APPLICATION NUMBER: 60/097,917
PRIOR FILING DATE: 1998-08-25
PRIOR APPLICATION NUMBER: 60/098,634
PRIOR FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 170
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO: 45
LENGTH: 2622
TYPE: DNA
ORGANISM: Homo sapiens
US-09-904-615-45

Query Match 34.4%; Score 51.6; DB 4; Length 2622;
Best Local Similarity 67.9%; Pred. No. 3.2e-08;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1 GAGATAGGAAAAACCGCTTAGGCTGAGAGTGGACCTGGGCGACATATCTGCTT 60
Db 1358 GAGATAGGAAAAACCGCTTAGGCTGAGAGTGGACCTGGGCGACATATCTGCTT 1299
Qy 61 GTAAGCACTGAGATGTTTATGTATGATCATATCTTAAAGCAGAC 106
Db 1298 GTTGCTCTGCTAGTATGATATTTGTGTAAGTAACATTAATC 1253

RESULT 13
US-08-991-789A-240
Sequence 240, Application US/08991789A
Patent No. 6225054
GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
Smith, John M.
Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292

CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 240:

SEQUENCE CHARACTERISTICS:

LENGTH: 519 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

MOLECULE TYPE: linear
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:

ORGANISM: Homo sapiens

SEQUENCE DESCRIPTION: SEQ ID NO: 240:
US-08-991-789A-240

Query Match 26.1%; Score 39.2; DB 3; Length 519;
Best Local Similarity 58.6%; Pred. No. 0.00033;
Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 7 GGGAAAAACCGCTTAGGCTGAGAGTGGACCTGGGCGACATATCTTGTAAAG 66
Db 227 GAGAAAAACCGCTTAGGCTGAGAGTGGACCTGGGCGACATATCTTGTATG 286
Qy 67 CACTGAGATGTTATGTATGATATCTTAAAGCAGACACTTAATCTTACAT 122
Db 287 CTTTACTCCAGAGATGTTGGCGGAGGAAACATTAATCTTGCCTTACGTCACAT 342

RESULT 14
US-09-062-451-240
Sequence 240, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:
APPLICANT: Frudakis, Tony N.
Smith, John M.
Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 297

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/062,451
FILING DATE: 04-APR-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

INFORMATION FOR SEQ ID NO: 240:
SEQUENCE CHARACTERISTICS:

LENGTH: 519 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

MOLECULE TYPE: linear
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

US-09-062-451-240

Query Match 26.1%; Score 39.2; DB 4; Length 519;
Best Local Similarity 58.6%; Pred. No. 0.00033;
Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 7 GGGAAAAACCGCTTAGGCTGAGAGTGGACCTGGGCGAGCAATATCTTGTAAAG 66
Db 227 GAGAAAAACCGCTTAGGCTGAGAGTGGACCTGGGCGAGCAATATCTTGTATG 286
Qy 67 CACTGAGATGTTATGTATGATATCTTAAAGCAGACACTTAATCTTACAT 122
Db 287 CTTTACTCCAGAGATGTTGGCGGAGGAAACATTAATCTTGCCTTACGTCACAT 342

RESULT 15

US-09-598-326-240

; Sequence 240, Application US/09598326

; Patent No. 6423496

GENERAL INFORMATION:

; APPLICANT: Prudakis, Tony N.

; Smith, John M.

; Reed, Steven G.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF BREAST CANCER

; NUMBER OF SEQUENCES: 247

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed Intellectual Property Law Group PLLC

; STREET: 701 Fifth Avenue, Suite 6300

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/598,326

; FILING DATE: 20-Jun-2000

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Poter, Jane E.R.

; REGISTRATION NUMBER: 33,332

; REFERENCE/DOCKET NUMBER: 210121.419D1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 240:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 519 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; SEQUENCE DESCRIPTION: SEQ ID NO: 240:

; US-09-598-326-240

Query Match 26.1%; Score 39.2; DB 4; Length 519;

Best Local Similarity 58.6%; Pred. No. 0.00033;

Matches 68; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
QY      7 GGGAAAAACCGCCTTAGCGCTGAGGTGGACCTGCGGCGACGACATACCTGCTTTGTAAG 66
      |||||
Db      227 GAGAAAAACCACTTATGCGCGGAGCGAGCATGTGGCAGCATGCTGCTGTTATG 286
      |||||
QY      67 CACGAGATGTTTGTGTATGATATCTAAAGCAGCACTTAATCCTTACAT 122
      |||||
Db      287 CTTACTCCACAGATGTTTGGCGGAGGAAAACATTAATCTGGCTTACGTGCACAT 342
      |||||
```

Search completed: February 28, 2004, 01:09:43
Job time : 21.2216 secs

This Page Blank (uspo)

Db 9361 ATTGCTATGATGCAAGACCTTTGTTTAC 9390

RESULT 2

US-09-421-112-45
; Sequence 45, Application US/09421112
; Publication No. US20030082532A1
; GENERAL INFORMATION:
; APPLICANT: NI, Jian
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: TNFR Related Gene 12
; FILE REFERENCE: P490
; CURRENT APPLICATION NUMBER: US/09/421,112
; CURRENT FILING DATE: 1999-10-19
; EARLIER APPLICATION NUMBER: 60/104,950
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 35414
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-421-112-45

Query Match 98.9%; Score 148.4; DB 10; Length 35414;
Best Local Similarity 99.3%; Pred. No. 7.9e-39;
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGATAGGGAACCCGCTTAGGGCTGAGGTGGACCTGGCGGCGCAATACCTGCTT 60
Db 9241 GAGATAGGGAACCCGCTTAGGGCTGAGGTGGACCTGGCGGCGCAATACCTGCTT 9300
Qy 61 GTAAGACATGAGATGTTATGTTATGATGATATCTAAAGACAGCACTTAATCCTTAC 120
Db 9301 GTAAGACATGAGATGTTATGTTATGATGATATCTAAAGACAGCACTTAATCCTTAC 9360
Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 9361 ATTGCTATGATGCAAGACCTTTGTTTAC 9390

RESULT 3

US-09-764-891-8024
; Sequence 8024, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8024
; LENGTH: 32249
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8024

Query Match 97.9%; Score 146.8; DB 10; Length 32249;
Best Local Similarity 98.7%; Pred. No. 2.6e-38;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAGATAGGGAACCCGCTTAGGGCTGAGGTGGACCTGGCGGCGCAATACCTGCTT 60
Db 26531 GAGATAGGGAACCCGCTTAGGGCTGAGGTGGACCTGGCGGCGCAATACCTGCTT 26590
Qy 61 GTAAGACATGAGATGTTATGTTATGATGATATCTAAAGACAGCACTTAATCCTTAC 120
Db 26591 GTAAGACATGAGATGTTATGTTATGATGATATCTAAAGACAGCACTTAATCCTTAC 26650

Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 26651 ATTGCTATGATGCAAGACCTTTGTTTAC 26680

RESULT 4

US-09-984-271-20/c
; Sequence 20, Application US/09984271
; Publication No. US20030040088A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/984,271
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/482,273
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US99/15849
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: 60/092,921
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,922
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: 60/092,956
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 2525
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1354)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-984-271-20

Query Match 97.3%; Score 146; DB 10; Length 2525;
Best Local Similarity 95.3%; Pred. No. 1.4e-38;
Matches 143; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGATAGGGAACCCGCTTAGGGCTGAGGTGGACCTGGCGGCGCAATACCTGCTT 60
Db 1210 GAGATAGGGAACCCGCTTAGGGCTGAGGTGGACCTGGCGGCGCAATACCTGCTT 1151
Qy 61 GTAAGACATGAGATGTTATGTTATGATGATATCTAAAGACAGCACTTAATCCTTAC 120
Db 1150 GTAAGACATGAGATGTTATGTTATGATGATATCTAAAGACAGCACTTAATCCTTAC 1091
Qy 121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
Db 1090 MTTGTTATGATGCAAGACCTTTGTTTAC 1061

RESULT 5

US-10-242-535A-15659
; Sequence 15659, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28

```

; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15659
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Human
; US-10-242-535A-15659

Query Match          96.8%; Score 145.2; DB 15; Length 390;
Best Local Similarity 98.0%; Pred. No. 1e-38;
Matches 147; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1  GAGATAGGAAAAACCCGCTTAGGGCTGAGGTGGAGACCTGCGGCGAGCAATACCTGCTT 60
Db      75  GAGATTAGGAAAAACCCGCTTAGGGCTGAGGTGGAGACCTGCGGCGAGCAATACCTGCTT 134

Qy      61  GTAAGACCTGAGATGTTATGTATGATGATATCTAAAGACAGACCTTAATCCTTTAC 120
Db     135  GTAAGACCTGAGATGTTATGTATGATGATATCTAAAGACAGACCTTAATCCTTTAC 194

Qy     121  ATTGCTATGATGCAAGACCTTGTCTAC 150
Db     195  ATTGCTATGATGCAAGACCTTGTCTAC 224

RESULT 6
US-10-292-798-1369/C
; Sequence 1369, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABRARANT, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US/10/292,798
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1369
; LENGTH: 744802
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(744802)
; NAME/KEY: CDS
; LOCATION: (201)..(246)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25640)..(25677)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27078)..(27094)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (141192)..(141769)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159571)..(159606)
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (174525)..(174575)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (234891)..(235013)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (235514)..(235560)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (279677)..(279729)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (408660)..(409123)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (409204)..(409669)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (428381)..(428396)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (472204)..(472330)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (714252)..(714355)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (714447)..(714529)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (739794)..(739891)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (51812)..(51911)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (57122)..(57221)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (79368)..(79467)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (293951)..(294050)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (310089)..(310188)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (332935)..(332935)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (332992)..(332992)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (362002)..(362101)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (639781)..(639880)
; OTHER INFORMATION: a, t, c, g, unknown or other
; US-10-292-798-1369
```

```
Query Match      95.7%; Score 143.6; DB 15; Length 744802;
Best Local Similarity 97.3%; Pred. No. 1.5e-36;
Matches 146; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGACCTGCGGGCAGCAATACCTGTTT 60
        |||
DB      635834 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGACCTGCGGGCAGCAATACCTGTTT 635775

QY      61 GTAAGACCTGAGATGTTTATGTGTATGATATCTAAAGCAGACGACTTAATCCTTTAC 120
        |||
DB      635774 GTAAGACCTGAGATGTTTATGTGTATGATATCTAAAGCAGACGACTTAATCCTTTAC 635715

QY      121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
        |||
DB      635714 CTGTCTATGATGCAAGACCTTTGTTTCAC 635685

RESULT 7
US-10-027-632-101981
; Sequence 101981, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101981
; LENGTH: 2398
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101981

Query Match      94.7%; Score 142; DB 15; Length 2398;
Best Local Similarity 96.7%; Pred. No. 3e-37; 5; Indels 0; Gaps 0;
Matches 145; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGACCTGCGGGCAGCAATACCTGTTT 60
        |||
DB      2136 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGACCTGCGGGCAGCAATACCTGTTT 2195

QY      61 GTAAGACCTGAGATGTTTATGTGTATGATATCTAAAGCAGACGACTTAATCCTTTAC 120
        |||
DB      2196 GTAAGACCTGAGATGTTTATGTGTATGATATCTAAAGCAGACGACTTAATCCTTTAC 2255

QY      121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
        |||
DB      2256 CTGTCTATGATGCAAGACCTTTGTTTCAC 2285

RESULT 8
US-10-027-632-101982
; Sequence 101982, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101982

US-10-027-632-101982

Query Match      94.7%; Score 142; DB 15; Length 2398;
Best Local Similarity 96.7%; Pred. No. 3e-37; 5; Indels 0; Gaps 0;
Matches 145; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGACCTGCGGGCAGCAATACCTGTTT 60
        |||
DB      2136 GAGATAGGAAAAACCCCTTAGGCTGAGGTGGACCTGCGGGCAGCAATACCTGTTT 2195

QY      61 GTAAGACCTGAGATGTTTATGTGTATGATATCTAAAGCAGACGACTTAATCCTTTAC 120
        |||
DB      2196 GTAAGACCTGAGATGTTTATGTGTATGATATCTAAAGCAGACGACTTAATCCTTTAC 2255

QY      121 ATTGCTATGATGCAAGACCTTTGTTTCAC 150
        |||
DB      2256 CTGTCTATGATGCAAGACCTTTGTTTCAC 2285

RESULT 9
US-10-027-632-101983
; Sequence 101983, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101983
```

LENGTH: 2398
TYPE: DNA
ORGANISM: Human
US-10-027-632-101983

Query Match 94.7%; Score 142; DB 15; Length 2398;
Best Local Similarity 96.7%; Pred. No. 3e-37;
Matches 145; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 60
2136 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 2195
61 GTAAGGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
2196 GTAAGGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 2255
121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
2256 CTGCTATGATGCAAGACCTTTGTTTAC 2285

RESULT 10
US-10-233-958-17
Sequence 17, Application US/10233958
Publication No. US20040009468A1
GENERAL INFORMATION:
APPLICANT: Mach, Bernard
APPLICANT: Conrad, Bernard
TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
TITLE OF INVENTION: Provirus
FILE REFERENCE: 23135-504
CURRENT APPLICATION NUMBER: US/10/233,958
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: 60/316,513
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 60/316,522
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 17
LENGTH: 975
TYPE: DNA
ORGANISM: Human endogenous retrovirus
US-10-233-958-17

Query Match 93.6%; Score 140.4; DB 15; Length 975;
Best Local Similarity 96.0%; Pred. No. 6.6e-37;
Matches 144; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

1 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 60
562 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 621
61 GTAAGGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
622 GTAAGGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 681
121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
682 CTGCTATGATGCAAGACCTTTGTTTAC 711

RESULT 11
US-10-233-958-4
Sequence 4, Application US/10233958
Publication No. US20040009468A1
GENERAL INFORMATION:
APPLICANT: Mach, Bernard
APPLICANT: Conrad, Bernard
TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis

TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
TITLE OF INVENTION: Provirus
FILE REFERENCE: 23135-504
CURRENT APPLICATION NUMBER: US/10/233,958
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: 60/316,513
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 60/316,522
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 2689
TYPE: DNA
ORGANISM: Human endogenous retrovirus
US-10-233-958-4

Query Match 93.6%; Score 140.4; DB 15; Length 2689;
Best Local Similarity 96.0%; Pred. No. 1.1e-36;
Matches 144; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

1 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 60
2282 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 2341
61 GTAAGGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
2342 GTAAGGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 2401
121 ATTGCTATGATGCAAGACCTTTGTTTAC 150
2402 CTGCTATGATGCAAGACCTTTGTTTAC 2431

RESULT 12
US-10-233-958-26/c
Sequence 26, Application US/10233958
Publication No. US20040009468A1
GENERAL INFORMATION:
APPLICANT: Mach, Bernard
APPLICANT: Conrad, Bernard
TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
TITLE OF INVENTION: Provirus
FILE REFERENCE: 23135-504
CURRENT APPLICATION NUMBER: US/10/233,958
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: 60/316,513
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 60/316,522
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 26
LENGTH: 6020
TYPE: DNA
ORGANISM: Human endogenous retrovirus
US-10-233-958-26

Query Match 93.6%; Score 140.4; DB 15; Length 6020;
Best Local Similarity 96.0%; Pred. No. 1.6e-36;
Matches 144; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

1 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 60
2194 GAGTATAGGAAAAACCCGCTTAGGGCTGAGAGTGGAGACCTGCGGCGAGCAATACCTGCTT 2135
61 GTAAGGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 120
2134 GTAAGGACCTGAGATGTTTATGTATGATGATATCTAAAGACAGCACTTAATCCTTTAC 2075

Qy	121	ATTGCTATGATGCAAGACCTTGTTCAC	150
Db	2074	CTTGCTATGATGCAAGACCTTGTTCAC	2045

```

RESULT 13
US-10-233-958-42
: Sequence 42, Application US/10233958
: Publication No. US20040009468A1
: GENERAL INFORMATION:
: APPLICANT: Mach, Bernard
: TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
: TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
: TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
: FILE REFERENCE: 23135-504
: CURRENT APPLICATION NUMBER: US/10/0233,958
: CURRENT FILING DATE: 2002-09-03
: PRIOR APPLICATION NUMBER: 60/316,513
: PRIOR FILING DATE: 2001-08-31
: PRIOR APPLICATION NUMBER: 60/316,522
: PRIOR FILING DATE: 2001-08-31
: NUMBER OF SEQ ID NOS: 46
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 42
: LENGTH: 10569
: TYPE: DNA
: ORGANISM: Human endogenous retrovirus
US-10-233-958-42

```

	Query Match	Similarity	%	Score	Id	DB	Length
Best Local	Similarity	93.6%	140.4	15	10569		
Matches	144	Conservative	0	Mismatches	6	Indels	0
						Gaps	0
QY	1	GAGATAGGAAAAA	CGCCCTTAGGG	CTGGAGGTGGAC	CTCGGGCAGCAATATCTGCTTT	60	
DB	9589	GAGATAGGAAAAA	CTGCCTTAGGG	CTGGAGGTGGAC	ATGTTGGCAGCAATATCTGCTTT	9648	
QY	61	GTTAAAGCAC	TGAGATTTATGT	GTATGATATCTTAAAA	GACACGACTTAATCCTTTAC	120	
DB	9649	GTTAAAGCAT	TGAGATTTATGT	ATGATATCTTAAAA	GACACGACTTAATCCTTTAC	9708	
QY	121	ATTGCTATGAT	GCAAGACCTTT	CTTCAC	150		
DB	9709	CTTGCTATGAT	GCAAGACCTTT	GTTCAC	9738		

```

RESULT 14
US-10-067-514-1/C
Sequence 1, Application US/10067514
Publication NO. US20030054531A1
GENERAL INFORMATION:
APPLICANT: Gletsdorf, Solveig
APPLICANT: Jonsdottir, Sif
TITLE OF INVENTION: HUMAN STROKE GENE
FILE REFERENCE: 2345-2010-003
CURRENT APPLICATION NUMBER: US/10/067,514
CURRENT FILING DATE: 2002-02-04
PRIOR APPLICATION NUMBER: US 09/611/352
PRIOR FILING DATE: 2001-03-19
NUMBER OF SEQ ID NOS: 84
SOFTWARE: Faetscq for Windows Version 4.0.
SEQ ID NO 1
LENGTH: 1691139
TYPE: DNA
ORGANISM: Human
US-10-067-514-1

```

Query Match	93.6%	Score 140.4;	DB 14;	Length 1691139;
Best Local Similarity	96.0%	Pred. No. 2.7e-35;		
Matches 144;	Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0

Qy	1	GAGTATAGGAAAAACCGCCTTAAAGGCTGAGAGTGAGACTGCGGGACGACATACGTCTT	60
Db	1166118	GAGATAGGAAAAAACCCGCTTAAAGCTGAGAGTGAGACATGCGGGACGACATACGTCTT	1166058
Qy	61	GTAAGACCTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTAC	120
Db	1166058	GTAAGACCTTGAGATGTTTATGTGTATGCATATCTAAAGCAGCACTTAATCCTTAC	1165999
Qy	121	ATTGCTATGATGCAGAACCTTTGTTAC	150
Db	1165998	CTTGCTATGATGCAGAACCTTTGTTAC	1165969

```

RESULT 15
US-10-419-723-1/c
; Sequence 1, Application US/10419723
; Publication No. US20040014099A1
; GENERAL INFORMATION:
; APPLICANT: Greteardottir, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reynisdottir, Sigridur Th.
; APPLICANT: Thorleifsson, Guðmar
; APPLICANT: Gulcher, Jeffrey
; TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
; TITLE OF INVENTION: METHODS OF TREATMENT
; FILE REFERENCE: 2345.2010-005
; CURRENT APPLICATION NUMBER: US/10/419,723
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: 10/255,120
; PRIOR FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 10/067,514
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 09/811,352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-419-723-1

```

Query Match	93.6%	Score 140.4	DB 15	Length 1691139
Best Local Similarity	96.0%	Prod. No. 2.7e-35		
Matches 144	Conservative 0	Mismatches 6	Indels 0	Gaps 0
QY	1	GAGATAGGGAAAAACCGCCTTAGGGCTCGAGGTGGGACCTGGCGGACGATACCTGCTTT	60	
Db	1166118	GAGATAGGGAAAAACCGCCTTTAAGGCTGGAGGTGGGACATGGCGGACGATACCTGCTTT	1166055	
QY	61	GTAAGAAGCTGAGATGTTTATGTATGACATATCTAAAGACAGACACTTAATCCTTTAC	120	
Db	1166058	GTAAGAAGCTGAGATGTTTATGTATGACATATCTAAAGACAGACACTTAATCCTTTAC	1165995	
QY	121	ATTGCTATGATGCAAGACCTTGGTTCAC	150	
Db	1165998	CTTGCTATGATGCAAGACCTTGGTTCAC	1165969	

Search completed: February 28, 2004, 01:17:19
Job time : 71.3061 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 27, 2004, 22:35:13 ; Search time 382.694 Seconds
(without alignments)
8285.900 Million cell updates/sec

Title: US-10-016-604-5

Perfect score: 879

Sequence: 1 tggggggaagaaagcaagagag.....ccctctctctctctctata 879

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 233733 seqs, 180373377 residues

Total number of hits satisfying chosen parameters: 4707466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
3: /cgn2_6/prodata/2/pubpna/US05_PUBCOMB.seq:*
4: /cgn2_6/prodata/2/pubpna/US04_PUBCOMB.seq:*
5: /cgn2_6/prodata/2/pubpna/US03_PUBCOMB.seq:*
6: /cgn2_6/prodata/2/pubpna/US02_PUBCOMB.seq:*
7: /cgn2_6/prodata/2/pubpna/US01_PUBCOMB.seq:*
8: /cgn2_6/prodata/2/pubpna/US00_PUBCOMB.seq:*
9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
11: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
12: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
13: /cgn2_6/prodata/2/pubpna/US05_PUBCOMB.seq:*
14: /cgn2_6/prodata/2/pubpna/US04_PUBCOMB.seq:*
15: /cgn2_6/prodata/2/pubpna/US03_PUBCOMB.seq:*
16: /cgn2_6/prodata/2/pubpna/US02_PUBCOMB.seq:*
17: /cgn2_6/prodata/2/pubpna/US01_PUBCOMB.seq:*
18: /cgn2_6/prodata/2/pubpna/US00_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	856.6	97.5	32249	10	US-09-764-891-8024 Sequence 8024, Ap
2	845.4	96.2	35414	9	US-09-836-607-45 Sequence 45, Appl
3	845.4	96.2	35414	10	US-09-421-112-45 Sequence 45, Appl
4	805.6	91.6	1691139	14	US-10-067-514-1 Sequence 1, Appl
5	805.6	91.6	1691139	15	US-10-419-723-1 Sequence 1, Appl
6	797.6	90.7	975	15	US-10-233-958-15 Sequence 15, Appl
7	797.6	90.7	975	15	US-10-233-958-17 Sequence 17, Appl
8	797.6	90.7	975	15	US-10-233-958-2 Sequence 2, Appl
9	797.6	90.7	975	15	US-10-233-958-4 Sequence 4, Appl
10	796	90.6	6020	15	US-10-233-858-26 Sequence 26, Appl
11	796	90.6	6020	15	US-10-233-858-26 Sequence 26, Appl
12	794.4	90.4	969	15	US-10-233-958-18 Sequence 18, Appl
13	794.4	90.4	969	15	US-10-233-958-16 Sequence 16, Appl
14	792.8	90.2	1010	15	US-10-233-958-19 Sequence 19, Appl

C	16	789.2	89.8	744802	15	US-10-292-798-1369	Sequence 1369, Ap
	17	784	89.2	198285	9	US-09-880-107-3814	Sequence 3814, Ap
	18	782.4	89.0	9343	15	US-10-233-958-43	Sequence 43, Appl
	19	778.6	88.6	1574	15	US-10-242-355-1080	Sequence 1080, Ap
	20	764.8	87.0	1043	9	US-09-764-869-1485	Sequence 1485, Ap
	21	764.8	87.0	1043	14	US-10-091-504-1485	Sequence 1485, Ap
	22	764.8	87.0	1043	15	US-10-227-577-1485	Sequence 1485, Ap
	23	763.8	86.9	2189	15	US-10-027-632-101674	Sequence 101674, Ap
	24	763.2	86.8	922	9	US-09-764-869-1486	Sequence 1486, Ap
	25	763.2	86.8	922	14	US-10-091-504-1486	Sequence 1486, Ap
	26	763.2	86.8	922	15	US-10-227-577-1486	Sequence 1486, Ap
	27	748.8	85.2	54433	15	US-10-085-117-124	Sequence 124, Appl
	28	738	84.0	2398	15	US-10-027-632-101981	Sequence 101981, Ap
	29	738	84.0	2398	15	US-10-027-632-101982	Sequence 101982, Ap
	30	738	84.0	2398	15	US-10-027-632-101983	Sequence 101983, Ap
	31	735.2	83.6	176373	13	US-10-095-407-117	Sequence 117, Appl
	32	721.2	82.0	3112	15	US-10-027-632-114936	Sequence 114936, Ap
	33	719.8	81.9	1148	10	US-09-764-891-8343	Sequence 8343, Ap
	34	719.8	81.9	1148	14	US-10-091-572-632	Sequence 632, Appl
	35	717.4	81.6	235033	14	US-10-301-844-1	Sequence 1, Appl
	36	717.4	81.6	237326	14	US-10-301-844-2	Sequence 2, Appl
	37	696	79.2	2126	15	US-10-027-632-98430	Sequence 98430, A
	38	692.2	78.7	2525	10	US-09-984-271-20	Sequence 20, Appl
	39	688.6	78.3	2454	15	US-10-104-047-275	Sequence 275, Appl
	40	682	77.6	855	10	US-09-764-891-8343	Sequence 8343, Ap
	41	682	77.6	855	14	US-10-205-428-776	Sequence 776, Appl
	42	680.4	77.4	855	10	US-09-764-891-8342	Sequence 8342, Ap
	43	680.4	77.4	855	14	US-10-205-428-775	Sequence 775, Appl
	44	608.8	69.3	1039	15	US-10-027-632-119427	Sequence 119427, Ap
	45	608.8	69.3	1039	15	US-10-027-632-119428	Sequence 119428, Ap

ALIGNMENTS

RESULT 1
US-09-764-891-8024
Sequence 8024, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764, 891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8024
LENGTH: 32249
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-8024

Query Match 97.5% Score 856.6; DB 10; Length 32249;
Best Local Similarity 96.4%; Pred. No. 1.8e-278;
Matches 865; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Cy	1	TGTTGGGAAAAGCAAGAGATGATGTTGTTCTGTCTGTGTAAGAGATGATCAT	60
Db	25971	TGTTGGGAAAAGCAAGAGATGATGTTGTTCTGTCTGTGTAAGAGATGATCAT	26030
Cy	61	AGGAGATCTCAATTTGTTATGTTAAGAAAATTTCTTGCCTTGAGATTCGTTATC	120
Db	26031	AGGAGATCTCAATTTGTTATGTTAAGAAAATTTCTTGCCTTGAGATTCGTTATC	26090
Cy	121	TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGCAACAGGTTAA	180
Db	26091	TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGCAACAGGTTAA	26150
Cy	181	ATGATTTAAGGCGCGTCAGAGATGCTTTGTTTAAAGATGCTTGAAGGACATGCTC	240
Db	26151	ATGATTTAAGGCGCGTCAGAGATGCTTTGTTTAAAGATGCTTGAAGGACATGCTC	26210

QY 241 CTTAAGATCATCACCACTCCCTAATCTCAAGTACCAGGAGACAAAAACTGCGAAGG 300
DB 26211 CTTAAGATCATCACCACTCCCTAATCTCAAGTACCAGGAGACAAAACTGCGAAGG 26270
QY 301 CCCAGGAGACTCTGCTAGAGAAAGCAGGATTTGTCACGTTTCTCCCATGTGTAG 360
DB 26271 CCCAGGAGACTCTGCTAGAGAAAGCAGGATTTGTCACGTTTCTCCCATGTGTAG 26330
QY 361 CCGAAATATGCTGCTGAGGAGAGGAGAACCGACCGTCCCGACCGGACCCGTA 420
DB 26331 TCTGAATATGCTGCTGAGGAGAGGAGAACCGACCGTCCCGACCGGACCCGTA 26390
QY 421 AAGGCTCTGCTGAGGAGAGATTAAGTAAAGAGAGAGATGCTCTTTCAGTTGAGACA 480
DB 26391 AAGGCTCTGCTGAGGAGAGATTAAGTAAAGAGAGAGATGCTCTTTCAGTTGAGACA 26450
QY 481 AGAGAAAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 26451 AGAGAAAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26510
QY 541 GATTGATGCTCATCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 26511 GATTGATGCTCATCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 26570
QY 601 GCGGAGAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 26571 GCGGAGAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26630
QY 661 CACAGCACTTAATCTTATCAATGCTGTATGATGCAAAAGCCTTGTTCACATGTTGCT 720
DB 26631 CACAGCACTTAATCTTATCAATGCTGTATGATGCAAAAGCCTTGTTCACATGTTGCT 26690
QY 721 GCTGACCTCTCCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 26691 GCTGACCTCTCCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26750
QY 781 CCACAGATGATCAATAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 26751 CCACAGATGATCAATAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 26810
QY 841 AACGCTGCTCCCGGCTCCCTTCTTCTTCTCTATA 879
DB 26811 AACGCTGCTCCCGGCTCCCTTCTTCTTCTCTATA 26849

RESULT 2
US-09-836-607-45
; Sequence 45, Application US/09836607
; Patent No. US20020098541A1
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: TNFR Related Gene 12
; FILE REFERENCE: PF490P1
; CURRENT APPLICATION NUMBER: US/09/836,607
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 60/198,388
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 09/421,112
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/104,950
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 45
; LENGTH: 35414
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-836-607-45

Query Match 96.2%; Score 845.4; DB 9; Length 35414;
Best Local Similarity 97.6%; Pred. No. 1,26-274;

Matches 858; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAGCAAGAGATCATGATTTGTTACTGTGCTGTGTAGAAAAGATGACAT 60
DB 8681 TGTGGGAAAAGCAAGAGATCATGATTTGTTACTGTGCTGTGTAGAAAAGATGACAT 8740
QY 61 AGGAGATCTCAATTTGTATGTAATAAGAAAATCTTCTGCTTGAAGATTCTGTTATC 120
DB 8741 AGGAGATCTCAATTTGTATGTAATAAGAAAATCTTCTGCTTGAAGATTCTGTTATC 8800
QY 121 TATGACCTTAACCCCAACCCCGTCTGTGAACAATGCTGTGTCACTCAGGGTTA 180
DB 8801 TATGACCTTAACCCCAACCCCGTCTGTGAACAATGCTGTGTCACTCAGGGTTA 8860
QY 181 ATGATTAAGGGGGGTGAGAGATGCTCTTGTAAACAGATGCTTGAAGCAGATGCTC 240
DB 8861 ATGATTAAGGGGGGTGAGAGATGCTCTTGTAAACAGATGCTTGAAGCAGATGCTC 8920
QY 241 CTTAAGATCATCACCACTCCCTAATCTCAAGTACCAGGAGACAAAACTGCGAAGG 300
DB 8921 CTTAAGATCATCACCACTCCCTAATCTCAAGTACCAGGAGACAAAACTGCGAAGG 8980
QY 301 CCCAGGAGACTCTGCTAGAGAAAGCAGGATTTGTCACGTTTCTCCCATGTGTAG 360
DB 8981 CCCAGGAGACTCTGCTAGAGAAAGCAGGATTTGTCACGTTTCTCCCATGTGTAG 9040
QY 361 CCGAATATGAGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 9041 CCGAATATGAGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9100
QY 421 AAGGCTCTGCTGAGAGAGATTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 9101 AAGGCTCTGCTGAGAGAGATTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAG 9160
QY 481 AGAGAAAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 9161 AGAGAAAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9220
QY 541 GATTGATGCTCATCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 9221 GATTGATGCTCATCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9280
QY 601 GCGGAGAGCAATCTGCTTGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 9281 GCGGAGAGCAATCTGCTTGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9340
QY 661 CACAGCACTTAATCTTATCAATGCTGTATGATGCAAAAGCCTTGTTCACATGTTGCT 720
DB 9341 CACAGCACTTAATCTTATCAATGCTGTATGATGCAAAAGCCTTGTTCACATGTTGCT 9400
QY 721 GCTGACCTCTCCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 9401 GCTGACCTCTCCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9460
QY 781 CCACAGATGATCAATAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 9461 CCACAGATGATCAATAATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9520
QY 841 AACGCTGCTCCCGGCTCCCTTCTTCTTCTCTATA 879
DB 9521 AACGCTGCTCCCGGCTCCCTTCTTCTTCTCTATA 9559

RESULT 3
US-09-421-112-45
; Sequence 45, Application US/09421112
; Publication No. US20030082532A1
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: TNFR Related Gene 12
; FILE REFERENCE: PF490
; CURRENT APPLICATION NUMBER: US/09/421,112

;; CURRENT FILING DATE: 1999-10-19
;; EARLIER APPLICATION NUMBER: 60/104,950
;; EARLIER FILING DATE: 1998-10-20
;; NUMBER OF SEQ ID NOS: 47
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 45
;; LENGTH: 35414
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-421-112-45

Query Match 96.2%; Score 845.4; DB 10; Length 35414;
Best Local Similarity 97.6%; Pred. No. 1.2e-274;
Matches 858; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTAGAAAAGAGATGACAT 60
DB 8681 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTAGAAAAGAGATGACAT 8740
QY 61 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTTCTGCTTGAAGATTCTGTATC 120
DB 8741 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTTCTGCTTGAAGATTCTGTATC 8800
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAAATGCTGTGTCTCACTCAGGGTTAA 180
DB 8801 TATGACCTTACCCCAACCCCGTCTCTGAAAATGCTGTGTCTCACTCAGGGTTAA 8860
QY 181 ATGATTTAAGGCGGTGAGAGATGTCTTTGTAAAGATGCTTGAAGGAGCATGCTC 240
DB 8861 ATGATTTAAGGCGGTGAGAGATGTCTTTGTAAAGATGCTTGAAGGAGCATGCTC 8920
QY 241 CTTAAGAGTATCAACAATCTCAAGTACCAAGGAGCAAAAATCTGGGAGG 300
DB 8921 CTTAAGAGTATCAACAATCTCAAGTACCAAGGAGCAAAAATCTGGGAGG 8980
QY 301 CCGCAGGAGCTCTGCTAGAAAAGCAGATTTGTCAAAGTTCTCCCATGTGATAG 360
DB 8981 CCGCAGGAGCTCTGCTAGAAAAGCAGATTTGTCAAAGTTCTCCCATGTGATAG 9040
QY 361 CTTGAAATATGCGCTCTGTGGGAAAGGAAAGATCTGACCCGACCCGACACCCGTA 420
DB 9041 TCTGAAATATGCGCTCTGTGGGAAAGGAAAGATCTGACCCGACCCGACACCCGTA 9100
QY 421 AAGGCTGTGTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGAAGTGAACA 480
DB 9101 AAGGCTGTGTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGAAGTGAACA 9160
QY 481 AGAGAGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 9161 AGAGAGAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9220
QY 541 GATTGTATGCTCCATCTAGATAGAGAAAACCCCTTGAAGGCTGAGAGTGAACCT 600
DB 9221 GATTGTATGCTCCATCTAGATAGAGAAAACCCCTTGAAGGCTGAGAGTGAACCT 9280
QY 601 GCGGCGAGCAATCTGTCTTTGTAAGCATGAGATGTTATGCTATGCAATTAAG 660
DB 9281 GCGGCGAGCAATCTGTCTTTGTAAGCATGAGATGTTATGCTATGCAATTAAG 9340
QY 661 CACAGCACTTAATCTTACATTTGTATGATGCAAGACCTTGTTCACATGTTTGTCT 720
DB 9341 CACAGCACTTAATCTTACATTTGTATGATGCAAGACCTTGTTCACATGTTTGTCT 9400
QY 721 GCTGACCTCTCCCAATTTGTTGTGACCTGACATCCCTCTTGTGAGAAACAC 780
DB 9401 GCTGACCTCTCCCAATTTGTTGTGACCTGACATCCCTCTTGTGAGAAACAC 9460
QY 781 CCAAGATGATCTGTAATTAATTAAGGAACTCAGAGGCTGGGGATCTCCATATGCTA 840
DB 9461 CCAAGATGATCTGTAATTAATTAAGGAACTCAGAGGCTGGGGATCTCCATATGCTA 9520
QY 841 AACGCTGTCCCGGGTCCCTTCTTCTCTATTA 879
|||||
```

DB 9521 AACGCTGTCCCGGGTCCCTTCTTCTCTATTA 9559

RESULT 4
US-10-067-514-1/c
; Sequence 1, Application US/10067514
; Publication NO. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gensetec, Inc.
; APPLICANT: Janssen-Cilag, Sif
; APPLICANT: Keynote, Sif
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

Query Match 91.6%; Score 805.6; DB 14; Length 1691139;
Best Local Similarity 95.5%; Pred. No. 4.7e-260;
Matches 840; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

```
QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTAGAAAAGAGATGACAT 60
DB 116679 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTAGAAAAGAGATGACAT 116620
QY 61 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTTCTGCTTGAAGATTCTGTATC 120
DB 116681 AGGAGACTCCATTTTGTATGTACTAAGAAAATTTCTTCTGCTTGAAGATTCTGTATC 116560
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAAATGCTGTGTCTCACTCAGGGTTAA 179
DB 116691 TATGACCTTACCCCAACCCCGTCTCTGAAAATGCTGTGTCTCACTCAGGGTTAA 116500
QY 181 ATGATTTAAGGCGGTGAGAGATGTCTTTGTAAAGATGCTTGAAGGAGCATGCTC 239
DB 116699 ATGATTTAAGGCGGTGAGAGATGTCTTTGTAAAGATGCTTGAAGGAGCATGCTC 116440
QY 241 CTTAAGAGTATCAACAATCTCAAGTACCAAGGAGCAAAAATCTGGGAGG 299
DB 116709 CTTAAGAGTATCAACAATCTCAAGTACCAAGGAGCAAAAATCTGGGAGG 116380
QY 301 CCGCAGGAGCTCTGCTAGAAAAGCAGATTTGTCAAAGTTCTCCCATGTGATAG 359
DB 116719 CCGCAGGAGCTCTGCTAGAAAAGCAGATTTGTCAAAGTTCTCCCATGTGATAG 116320
QY 361 GCTGAAATATGCGCTCTGTGGGAAAGGAAAGATCTGACCCGACCCGACACCCGT 419
DB 116729 GCTGAAATATGCGCTCTGTGGGAAAGGAAAGATCTGACCCGACCCGACACCCGT 116260
QY 421 AAAAGGCTGTGTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGAAGTGAAC 479
DB 116739 AAAAGGCTGTGTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGAAGTGAAC 116200
QY 481 AAGAGAGAGCATTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
DB 116749 AAGAGAGAGCATTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 116140
QY 540 CGATTGATCTCCATCTAGATAGAGAAAACCCCTTGAAGGCTGAGAGTGAAC 599
DB 116759 CGATTGATCTCCATCTAGATAGAGAAAACCCCTTGAAGGCTGAGAGTGAAC 116080
QY 600 TCGGGGAGCAATCTGCTTTGTAAGCACTGAGATGTTATGCTATCTAATAA 659
DB 116769 TCGGGGAGCAATCTGCTTTGTAAGCACTGAGATGTTATGCTATCTAATAA 116020
```

QY	660	CCACAGACACCTTAATCCCTTTCATCATGTGTCTATATGCAAGAACCCTTGTGCATATGTTTGC	719
Db	1166019	GCAACAGACCTTGAATCTTCTTACCTGTCTATATGCAAGAACCCTTGTGCATGTTTGTGC	1165960
QY	720	TGCTGACCCCTCTCCCAACAATGTCTTGTGACCCCTGACACATCCCTCTTTCGAGAAACA	779
Db	1165959	TGCTGACCCCTCTCCCAACTATTTGCTGTGTGACCCCTGACACATCCCTCTTCAGAGAAACA	1165900
QY	780	CCACACAGATGTCAGTAAATACTAAGGGAACCTCAGAGGCTGGCGGGAATCCTTCATATGCT	839
Db	1165899	CCACACGATGATCAATAAATACTAAGGGAACCTCAGAGGCTGGCGGGAATCCTTCATATGCT	1165840
QY	840	GAACGCTGTGTCCTCCCGGGTCCCTCTCTTCTCTTCTCTATA	879
Db	1165839	GAACGCTGTGTCCTCCCGGGTCCCTCTATTTCTTCTCTATA	1165800

```

RESULT 5
US-10-419-723-1/c
; Sequence 1, Application US/10419723
; Publication No. US20040014099A1
; GENERAL INFORMATION:
; APPLICANT: Getarsdotcitr, Solveig
; APPLICANT: Jonadotcitr, Sif
; APPLICANT: Reymisdotcitr, Sigridur Th.
; APPLICANT: Thorleifsson, Gudmar
; APPLICANT: Gulcher, Jeffrey
; TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
; TITLE OF INVENTION: METHODS OF TREATMENT
; FILE REFERENCE: 2345.2010-005
; CURRENT APPLICATION NUMBER: US/10/419,723
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: 10/255,120
; PRIOR FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 10/067,514
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 09/811,352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-419-723-1

```

	Query Match	91.6%	Score 805.6	DB 15	Length 169139
	Best Local Similarity	95.5%	Prod. No. 4.7e-260		
	Matches	840	Conservative	0	Mismatches 33
				Indels	1
				Gaps	1
Qy	1	TGTGGGGAAAAGCAGAGAGATTCAGATTGTTACTGTGTCTGTGTGAGAAAGAGTAGACAT	60		
Db	1166679	TGTGGGGAAAAGCAGAGAGATTCAGATTGTCATCTGTGTGTGTGAGAAAGAGTAGACAT	1166620		
Qy	61	AGGAGACTCCATTTTGTATTGTACTAGAAAAATTTCTGTGCTTGAGATTCTGTATC	120		
Db	1166619	AGGAGACTCCATTTTGTGTCTGACACTAGAAAAATTTCTGTGAGATTCTGTATC	1166560		
Qy	121	TATGACCTTACCCCCAACCCCGTGCTCTGAAACATGTGCTGTGTC-CACTCAGGGTTA	179		
Db	1166559	TGTGACCTTACCCCCAACCCCGTGCTCTCTGAAATATGTGCTGTGTCAAATCTCAGGGTTG	1166500		
Qy	180	AATGAGATTAAAGGGGGGTGCAGAGATGTGTTTAAACAGATGCTTGAAGGACAGATGCT	239		
Db	1166499	AATGAGATTAAAGGGTTGTGCAGAGATGTGCTTTGTTAAACAGATGCTTGAAGGACAGATGCT	1166440		
Qy	240	CCTTAAAGTCATCACCACTCCCTATATCTCAAGTACCCAGGAGACAAAACTGGGAAAG	299		
Db	1166439	CCTTAAAGTCATCACCACTCCCTATATCTCAAGTACCCAGGAGACAAAACTGGGAAAG	1166380		
Qy	300	GCOCAGAGGACCTCTGCCTAGAGAAAGCAGAGATTGTCCAAAGTTTCCGCCCATGTGTA	359		
Db	1166379	GCTGACAGGACCTCTGCCTAGAGAAAGCAGAGATTGTTCAAAGTTTCTCCCATGTGTA	1166320		

Oy	360	GCCTGAATATTTGGCTCTGATGGAAAGGAAAGACCTGACCGCTCCCGACGCCGACAACCGGT	419
Db	1166319	GTCGTAAATATTTGGCTCTGATGGAAAGGAAAGACCTGACCGCTCCCGACGCCGACAACCGGT	1166260
Oy	420	AAAGGCTCTGTGCTGAGAGAGATTAAGTAAAGAGAGAAATGCTCTTTCAGTTGAGAC	479
Db	1166259	AAAGGCTCTGTGCTGAGAGAGATTAAGTAAAGAGAAATGCTCTTTCAGTTGAGAC	1166200
Oy	480	AAGAGAAAGGCATCTGTCTCTCTGCTGTCCCTGGGCATGGAATGTTCCGGTATAAACC	539
Db	1166199	AAGAGAAAGGCATCTGTCTCTCTGCTCCCTCCCTGGGCATGGAATGTTCTCATATAAAC	1166140
Oy	540	CGATTGTATCTCCATCTACTGAGATTTGGGAAAAACCGCTTTAGGGCTGGAGGGGAGCC	599
Db	1166139	CGATTGTATCTCCATCTACTGAGATTTGGGAAAAACCGCTTTAGGGCTGGAGGGGAGCA	1166080
Oy	600	TGCGGGAGAGAATTAAGTCTTTGTTAAAGCACTGAGATGTTATGTATGATATCTTAA	659
Db	1166079	TGCGGGAGAGAATTAAGTCTTTGTTAAAGCACTTGAATGTTATGTATGATATCTTAA	1166020
Oy	660	GCAAGACACTTAAATCCCTTTACATTTGCTATGATGCAAAAGACCTTTGTTCAATGTTTGC	719
Db	1166019	GCAAGACACTTAAATCCCTTTACATTTGCTATGATGCAAAAGACCTTTGTTTCAAGTTTGC	1165960
Oy	720	TGCTGACCCCTCTCCGCCAATTTGCTTTGTACACCTGACACATCCCTCTTTCGAGAAACA	779
Db	1165959	TGCTGACCCCTCTCCGCCAATTTGCTTTGTACACCTGACACATCCCTCTTTCGAGAAACA	1165900
Oy	780	CCGACAGATATGCTAATAATTAAGAGAACTGAGAGGCTGGGGGAAATCCCTCATATGCT	839
Db	1165899	CCGACAGATATGCTAATAATTAAGAGAACTGAGAGGCTGGGGGAAATCCCTCATATGCT	1165840
Oy	840	GAACGCTGTTCCCGGGGCTCCCTCTTTCTTTCTATTA	879
Db	1165839	GAACGCTGTTCCCGGGGCTCCCTATTTCTTTCTATTA	1165800

```

RESULT 6
US-10-233-958-15
; Sequence 15, Application US/102339958
; Publication No. US20040009466A1
GENERAL INFORMATION:
APPLICANT: Mach, Bernard
TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
TITLE OF INVENTION: Theoret and Use in the Determination of Genetic
TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
FILE REFERENCE: 23135-504
CURRENT APPLICATION NUMBER: US/10/233,958
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: 60/316,513
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 60/316,522
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentm Ver. 2.1
SEQ ID NO 15
LENGTH: 975
TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-15

Query Match          90.7%; Score 797.6; DB 15; Length 975;
Best Local Similarity 94.9%; Pred. No. 2e-259;
Matches 835; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

Qy      1 TGTGGGAAAAAGCAGAGAGATCGATTGTTACTGTGTCTGTGTAGAAAGAATAGACAT 60
        |||||
Db      1 TGTGGGAAAAAGCAGAGAGATCGATTGTTACTGTGTCTGTGTAGAAAGAATAGACAT 60

Qy      61 AGGAGACTCCATTGTTATGTACTAGAAAAAATCTTCCTGCCCTTGAGATTCTGTTATC 120

```

```

Db 61 AGGAGACTCCATTTTGTCTGTACTAGAAAAATTTCTCTCCGTGAGATCTGTATATC 120
Qy 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTC-CACTCAGGGTTA 179
Db 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTC-CACTCAGGGTTA 180
Qy 180 AATGATTTAAGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGACATGCT 239
Db 181 AATGATTTAAGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGACATGCT 240
Qy 240 CTTTAAAGTATATCACTCTCTTAATCTCAAGTACCAGGACACAAAACCTCGGAAG 299
Db 241 CATTAAAGTATATCACTCTCTTAATCTCAAGTACCAGGACACAAAACCTCGGAAG 300
Qy 300 GCCGCAAGGACCTCTGCTAGGAAAGCCAGATATTTGCCACGTTTCTCCCATGTGATA 359
Db 301 GCCGCAAGGACCTCTGCTAGGAAAGCCAGATATTTGCCACGTTTCTCCCATGTGATA 360
Qy 360 GCCTGAATATAGCCTCGTGGGAAAGGAAAGACCTGACCCGCCCCAGACCCGCT 419
Db 361 GTCTGAATATAGCCTCGTGGGAAAGGAAAGACCTGACCAATCCCGACCAACCCGT 420
Qy 420 AAAGGCTCTGTCTGAGAGGATTAGTAAAGAGAAAGAAATGCTCTTGACGTTGAGAC 479
Db 421 AAAGGCTCTGTCTGAGAGGATTAGTAAAGAGAAAGAAATGCTCTTGACGTTGAGAC 480
Qy 480 AAGAGGAAGCATCTGTCTCTGCTGCTCCCTGGGCAATGGAATGTCTCGGTATATAAAC 539
Db 481 AAGAGGAAGCATCTGTCTCTGCTGCTCCCTGGGCAATGGAATGTCTCGGTATATAAAC 540
Qy 540 CGATTGTATGCTCCATCTACTAGATAGGAAAGAAACCGCTTAGAGGCTGAGGTGGAGC 599
Db 541 CGATTGTATGCTCCATCTACTAGATAGGAAAGAAACCGCTTAGAGGCTGAGGTGGAGC 600
Qy 600 TGGCGGAGCAATATCTGCTTTGTAAGACACTGAGATGTTATGTGATGCAATCTAAAA 659
Db 601 TGTGGGAGCAATATCTGCTTTGTAAGACACTGAGATGTTATGTGATGCAATCTAAAA 660
Qy 660 GCACAGACATTAATCTTTACATGTGTATGATGCAAGAACCTTTGTTACATGTTTGTG 719
Db 661 GCACAGACATTAATCTTTACATGTGTGTATGATGCAAGAACCTTTGTTACATGTTTGTG 720
Qy 720 TGTGACCTCTCTCCCAATATGCTGTGACCTGACCAATGCCCTTTGAGAAACA 779
Db 721 TGTGACCTCTCTCCCAATATGCTGTGATGCTGTGACCTGACCAATGCCCTTTGAGAAACA 780
Qy 780 CCCACAGATGATCAATTAATCTAAAGGAACTCAGAGGCTGGCGGATCTCCATATGCT 839
Db 781 CCCACAGATGATCAATTAATCTAAAGGAACTCAGAGGCTGGCGGATCTCCATATGCT 840
Qy 840 GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTCTATA 879
Db 841 GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTCTATA 880

```

```

RESULT 7
US-10-233-958-17
; Sequence 17, Application US/10233958
; Publication No. US2004009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: Thereof and Use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; PRIOR FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522

```

```

; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 17
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-17

Query Match      90.7%; Score 797.6; DB 15; Length 975;
Best Local Similarity 94.9%; Pred. No. 2e-259;
Matches 835; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

Qy 1 TGTGGGAAAAAGCAAGAGATCAGATTGTTACTGTGTGTGTAAGAAAGATAGACAT 60
Db 1 TGTGGGAAAAAGCAAGAGATGAGATTGTTACTGTGTGTGTATAGAAAGATAGACAT 60
Qy 61 AGGAGACTCCATTTTGTATGTACTAAGAAAAATTTCTGCTTGAATCTGTATATC 120
Db 61 AGGAGACTCCATTTTGTATGTACTAAGAAAAATTTCTGCTTGAATCTGTATATC 120
Qy 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTC-CACTCAGGGTTA 179
Db 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTGTC-CACTCAGGGTTA 180
Qy 180 AATGATTTAAGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGACATGCT 239
Db 181 AATGATTTAAGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGACATGCT 240
Qy 240 CTTTAAAGTATATCACTCTCTTAATCTCAAGTACCAGGACACAAAACCTCGGAAG 299
Db 241 CATTAAAGTATATCACTCTCTTAATCTCAAGTACCAGGACACAAAACCTCGGAAG 300
Qy 300 GCCGCAAGGACCTCTGCTAGGAAAGCCAGATATTTGCCACGTTTCTCCCATGTGATA 359
Db 301 ACCGCAAGGACCTCTGCTAGGAAAGCCAGATATTTGCCACGTTTCTCCCATGTGATA 360
Qy 360 GCCTGAATATAGCCTCGTGGGAAAGGAAAGACCTGACCCGCCCCAGACCCGCT 419
Db 361 GTCTGAATATAGCCTCGTGGGAAAGGAAAGACCTGACCCGCCCCAGACCCGCT 420
Qy 420 AAAGGCTCTGTCTGAGAGGATTAGTAAAGAGAAAGAAATGCTCTTGACGTTGAGAC 479
Db 421 AAAGGCTCTGTCTGAGAGGATTAGTAAAGAGAAAGAAATGCTCTTGACGTTGAGAC 480
Qy 480 AAGAGGAAGCATCTGTCTCTGCTGCTCCCTGGGCAATGGAATGTCTCGGTATATAAAC 539
Db 481 AAGAGGAAGCATCTGTCTCTGCTGCTCCCTGGGCAATGGAATGTCTCGGTATATAAAC 540
Qy 540 CGATTGTATGCTCCATCTACTAGATAGGAAAGAAACCGCTTAGAGGCTGAGGTGGAGC 599
Db 541 CGATTGTATGCTCCATCTACTAGATAGGAAAGAAACCGCTTAGAGGCTGAGGTGGAGC 600
Qy 600 TGGCGGAGCAATATCTGCTTTGTAAGACACTGAGATGTTATGTGATGCAATCTAAAA 659
Db 601 TGTGGGAGCAATATCTGCTTTGTAAGACACTGAGATGTTATGTGATGCAATCTAAAA 660
Qy 660 GCACAGACATTAATCTTTACATGTGTATGATGCAAGAACCTTTGTTACATGTTTGTG 719
Db 661 GCACAGACATTAATCTTTACATGTGTATGATGCAAGAACCTTTGTTACATGTTTGTG 720
Qy 720 TGTGACCTCTCTCCCAATATGCTGTGACCTGACCAATGCCCTTTGAGAAACA 779
Db 721 TGTGACCTCTCTCCCAATATGCTGTGATGCTGTGACCTGACCAATGCCCTTTGAGAAACA 780
Qy 780 CCCACAGATGATCAATTAATCTAAAGGAACTCAGAGGCTGGCGGATCTCCATATGCT 839
Db 781 CCCACAGATGATCAATTAATCTAAAGGAACTCAGAGGCTGGCGGATCTCCATATGCT 840
Qy 840 GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTCTATA 879
Db 841 GAAAGCTGTGTTCCCGGGTCCCTTTCTTTCTCTATA 880

```


Qy	360	CCCGAAATATGAGCCTCGTGGGAAAGGAAAGACTCGACCGTCCCCACCCGACACCGGT	419
Db	2081	GTCGAAATATGAGCCTCGTGGGAAAGGAAAGACTCGACCATCCCCAACCAACCCGT	2140
Qy	420	AAAGGCTCTGCTCAGAGAGATTTAGTAAAGAGAGAGAAATGCTCTTGGAGTTGAGAC	479
Db	2141	AAAGGCTCTGCTCAGAGAGATTTAGTAAAGAGAGAGAAATGCTCTTGGAGTTGAGAG	2200
Qy	480	AAAGAGAAAGCATCTGTCTCCTGCTGTCCTCGGGCAAATGAAATGCTCGATATAAAC	539
Db	2201	AAAGAGAAAGCATCTGTCTCCTGCTGTCCTCGGGCAAATGAAATGCTCAGATATAAAC	2260
Qy	540	CGATTGTATGCTCCATCTACTGAGATAGGAGAAAAACCCCTTAGGGCTGAGGTGGAC	599
Db	2261	CGATTGTAACTTCCATCTACTGAGATAGGAGAAAAACCTCCTTAGGGCTGAGGTGGACA	2320
Qy	600	TGCGGGGAGCAATCTGCTTTGTAAAGACATGAGATGTTATGTGTATGACATCTAA	659
Db	2321	TGTGGGAGCAATCTGCTTTGTAAAGACATGAGATGTTATGTGTATGACATCTAA	2380
Qy	660	GCACAGCACTTAATCTTTACATTTGCTATGATGCAAGACCTTGTTCACATGTTTGC	719
Db	2381	GCACAGCACTTGATCTTTACTCTTCTATGATGCAAGACCTTGTTCACATGTTTGC	2440
Qy	720	TGCTGACCCCTCTCCCAACAATTTGCTTGACCTTGACACATCCCCCTTTCGAGAAACA	779
Db	2441	TGCTGACCCCTCTCCCAACAATTTGCTTGACCTTGACACATCCCCCTCTCGAGAAACA	2500
Qy	780	CCCACAGATGATCACTAAATACTAAGGAACTCAGAGGCTGGCGGGATCTTCATATGCT	839
Db	2501	CCCACAGATGATCAATAAATACTAAGGAACTCAGAGGCTGGCGGGATCTTCATATGCT	2560
Qy	840	GAGCGTGGTCCCGGGGTCCCTCTTCTTCTCTATA 879	
Db	2561	GAGCGTGGTCCCGGGGCCCTTATTTCTTCTATA 2600	

```

RESULT 10
US-10-233-958-20
; Sequence 20, Application US/102339958
; Publication No. US20040009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Allelic Variance of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: Theeetof and Use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; CURRENT FILING DATE: 2002-08-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-20

Query Match      90.6%; Score 796; DB 15; Length 972;
Best Local Similarity 94.8%; Pred. No. 6,9e-259;
Matches 834; Conservative 0; Mismatches 45; Indels 1; Gaps 1

OY      1  TGTGGGGAAGCAAGAGAGATCAATTGTTACTGTGTGTGTGAAGAAGTAGACAT 60
Db      1  TGTGGGGAAGCAAGAGAGAGATCAATTGTTACTGTGTGTGTGAAGAAGTAGACAT 60
OY      61 AGGAGACTTCATTTGTTATGTACTAAGAAAAATTCCTTGCCTTGAGATTCTGTTATC 120
Db      61 AGGAGACTTCATTTGTTATGTACTAAGAAAAATTCCTTGCCTTGAGATTCTGTTATC 120

```

[illegible]

```

RESULT 11
US-10-233-958-26/c
; Sequence 26, Application US/10233958
; Publication No. US20040009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: Theoret and use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; TITLE OF INVENTION: Provirus
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522
; PRIOR FILING DATE: 2001-08-31

```

NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 26
LENGTH: 6020
TYPE: DNA
ORGANISM: Human endogenous retrovirus
US-10-233-958-26

Query Match 90.6%; Score 796; DB 15; Length 6020;
Best Local Similarity 94.8%; Pred. No. 2.2e-258;
Matches 834; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

```
QY 1 TGTGGGAAAAGCAAGAGATCAATTTGTTCTGTGTGTAAGAAAGTAGACAT 60
DB 2755 TGTGGGAAAAGCAAGAGATGAGATTGTTACTGTGTGTATAGAAAGTAGACAT 2696
QY 61 AGGAGACTCCATTTGTTGTTACTAAGAAAATTTCTGCGCTGAGATTGTTATC 120
DB 2695 AGGAGACTCCATTTGTTGTTACTAAGAAAATTTCTGCGCTGAGATTGTTATC 2636
QY 121 TATGACCTTACCCCAACCCCGTCTCTGMAACATGCTGTGTC-CACTCAGGGTTA 179
DB 2635 TATGACCTTACCCCAACCCCGTCTCTGMAACATGCTGTGTC-CACTCAGGGTTA 2576
QY 180 AATGATTTAAGGCGGTGCAAGATGTGTTTAAACAGATGCTTGAAGGAGATGCT 239
DB 2575 AATGATTTAAGGCGGTGCAAGATGTGTTTAAACAGATGCTTGAAGGAGATGCT 2516
QY 240 CCTTAAGATGATCAGCACTCCCTAATCTCAAGTACCAGGAGACAAAACCTGGGAG 299
DB 2515 CATTAAAGTATACACACTCCCTAATCTCAAGTACCAGGAGACAAAACCTGGGAG 2456
QY 300 GCCGCAAGGACCTTGTCTAGAAAGCAGGTATTTGTCACAGTTTCTCCCATGTGATA 359
DB 2455 ACCGCAAGGACCTTGTCTAGAAAGCAGGTATTTGTCACAGTTTCTCCCATGTGATA 2396
QY 360 GCCTGAATATGCGCTGTGTGGAAGGAAAGACCTGACCGTCCCGCAGCCGACACCGT 419
DB 2395 GTCGAAATATGCGCTGTGTGGAAGGAAAGACCTGACCGTCCCGCAGCCGACACCGT 2336
QY 420 AAGGCGTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTGTGAGTTGAGAC 479
DB 2335 AAGGCGTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTGTGAGTTGAGAC 2276
QY 480 AAGAGAAAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
DB 2275 AAGAGAAAGCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2216
QY 540 CGATTGATGCTCATCTACTAGATAGGAAAGACCGCTTACGCGTGAAGGTGGAGC 599
DB 2215 CGATTGATGCTCATCTACTAGATAGGAAAGACCGCTTACGCGTGAAGGTGGAGC 2156
QY 600 TGGGGGAGCAATCTGCTTGTAAAGCATGAGATGTTATGTTATGATATCTTAA 659
DB 2155 TGGGGGAGCAATCTGCTTGTAAAGCATGAGATGTTATGTTATGATATCTTAA 2096
QY 660 GCAAGACATTAATCTTTCATTTGCTATGATGCAAGACCTTGTGTCATGTTTGTG 719
DB 2095 GCAAGACATTTGATCTTTTACCTTGTGTATGATGCAAGACCTTGTGTCATGTTTGTG 2036
QY 720 TGTGACCTCTCTCCCAATTTGTTGACCTGACACATCCCTCTTTCGAGAAACA 779
DB 2035 TGTGACCTCTCTCCCAATTTGTTGACCTGACACATCCCTCTTTCGAGAAACA 1976
QY 780 CCCGCAATGATCTAGTAAATCTAAGGAACTCAGAGGCTGGGAGATCCCTCAATAGCT 839
DB 1975 CCCGCAATGATCTAATAATCTAAGGAACTCAGAGGCTGGGAGATCCCTCAATAGCT 1916
QY 840 GAAAGCTGTGTTCCCGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 879
DB 1915 GAAAGCTGTGTTCCCGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1876
```

RESULT 12
US-10-233-958-42
Sequence 42; Application US/10233958
Publication No. US2004009468A1
GENERAL INFORMATION:
APPLICANT: Mach, Bernard
TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
TITLE OF INVENTION: thereof and Use in the Determination of Genetic
TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
FILE REFERENCE: 23135-504
CURRENT APPLICATION NUMBER: US/10/233,958
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: 60/316,513
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 60/316,522
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
LENGTH: 10569
SEQ ID NO 42
TYPE: DNA
ORGANISM: Human endogenous retrovirus
US-10-233-958-42

Query Match 90.6%; Score 796; DB 15; Length 10569;
Best Local Similarity 94.8%; Pred. No. 3.2e-258;
Matches 834; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

```
QY 1 TGTGGGAAAAGCAAGAGATCAATTTGTTCTGTGTGTAAGAAAGTAGACAT 60
DB 765 TGTGGGAAAAGCAAGAGATGAGATTGTTACTGTGTGTATAGAAAGTAGACAT 824
QY 61 AGGAGACTCCATTTGTTGTTACTAAGAAAATTTCTGCGCTGAGATTGTTATC 120
DB 825 AGGAGACTCCATTTGTTGTTACTAAGAAAATTTCTGCGCTGAGATTGTTATC 884
QY 121 TATGACCTTACCCCAACCCCGTCTCTGMAACATGCTGTGTC-CACTCAGGGTTA 179
DB 885 TATGACCTTACCCCAACCCCGTCTCTGMAACATGCTGTGTC-CACTCAGGGTTA 944
QY 180 AATGATTTAAGGCGGTGCAAGATGTGTTTAAACAGATGCTTGAAGGAGATGCT 239
DB 945 AATGATTTAAGGCGGTGCAAGATGTGTTTAAACAGATGCTTGAAGGAGATGCT 1004
QY 240 CCTTAAGATGATCAGCACTCCCTAATCTCAAGTACCAGGAGACAAAACCTGGGAG 299
DB 1005 CATTAAAGTATACACACTCCCTAATCTCAAGTACCAGGAGACAAAACCTGGGAG 1064
QY 300 GCCGCAAGGACCTTGTCTAGAAAGCAGGTATTTGTCACAGTTTCTCCCATGTGATA 359
DB 1065 CCGGCAAGGACCTTGTCTAGAAAGCAGGTATTTGTCACAGTTTCTCCCATGTGATA 1124
QY 360 GCCTGAATATGCGCTGTGTGGAAGGAAAGACCTGACCGTCCCGCAGCCGACACCGT 419
DB 1125 GTCGAAATATGCGCTGTGTGGAAGGAAAGACCTGACCGTCCCGCAGCCGACACCGT 1184
QY 420 AAGGCGTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTGTGAGTTGAGAC 479
DB 1185 AAGGCGTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTGTGAGTTGAGAC 1244
QY 480 AAGAGAAAGCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
DB 1245 AAGAGAAAGCATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304
QY 540 CGATTGATGCTCATCTACTAGATAGGAAAGACCGCTTACGCGTGAAGGTGGAGC 599
DB 1305 CGATTGATGCTCATCTACTAGATAGGAAAGACCGCTTACGCGTGAAGGTGGAGC 1364
QY 600 TGGGGGAGCAATCTGCTTGTAAAGCATGAGATGTTATGTTATGATATCTTAA 659
DB 1365 TGGGGGAGCAATCTGCTTGTAAAGCATGAGATGTTATGTTATGATATCTTAA 1424
```


QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAAAGATGCTGTGTC -CACTCAGGGTTA 179
DB 121 TATGACCTTACCCCAACCCCGTCTCTGAAAAGATGCTGTGTC -CACTCAGGGTTA 180
QY 180 AATGAGTTAAGGGGGTGGCAGAGATGCTGTTTAAACAATGCTTAAAGGCAAGCT 239
DB 181 AATGAGTTAAGGGGGTGGCAGAGATGCTGTTTAAACAATGCTTAAAGGCAAGCT 240
QY 240 CCTTAAGATCATCACCACCTCCCTAATCTCAAGTACCAGGACACAAAAATCGCGAAG 299
DB 241 CATTTAAGATCATCACCACCTCCCTAATCTCAAGTACCAGGACACAAAAATCGCGAAG 300
QY 300 GCGCAGAGACCTCTGCTAGGAAAGCAGATATGTCACACGTTTCTCCCATGTGATA 359
DB 301 GCGCAGAGACCTCTGCTAGGAAAGCAGATATGTCACACGTTTCTCCCATGTGATA 360
QY 360 GCGTGAATATGCTGCTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACACCCGT 419
DB 361 GCTGAAATATGCTGCTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACACCCGT 420
QY 420 AAGAGGCTGCTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTGAGATTGAGAC 479
DB 421 AAGAGGCTGCTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTGAGATTGAGAC 480
QY 480 AAGAGAGGACATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
DB 481 AAGAGAGAGACATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 540 CGATTGATGCTCCATCTTACTGAGATAGGAAAAACCGCTTGAAGGCTGGAGTGGAGC 599
DB 541 CGATTGATGCTCCATCTTACTGAGATAGGAAAAACCGCTTGAAGGCTGGAGTGGAGC 600
QY 600 TGGGGGAGAGAACTCTGCTTGTAAAGACACTGAGATGTTATGTAAGCATATCTAAA 659
DB 601 TGTGGGAGAGAACTCTGCTTGTAAAGACACTGAGATGTTATGTAAGCATATCTAAA 660
QY 660 GCACAGCACTTAATCTCTTACATCTGTCTATGATGCAAAAGACCTTGTTCACATGTTGTC 719
DB 661 GCACAGCACTTAATCTCTTACATCTGTCTATGATGCAAAAGACCTTGTTCACATGTTGTC 720
QY 720 TGTGACCCCTCTCCCAACATGCTTGTGACCTGACACATCCCTCTTCGAGAAACA 779
DB 721 TGTGACCCCTCTCCCAACATGCTTGTGACCTGACACATCCCTCTTCGAGAAACA 780
QY 780 CCACAGATATATCTAATAATCTAAGGAACTGAGAGCTGGGGAGTCCCTCATATGCT 839
DB 781 CCACAGATATATCTAATAATCTAAGGAACTGAGAGCTGGGGAGTCCCTCATATGCT 840
QY 840 GAAAGCTGTGTTCCCGGGGCTCCCTTCTTTCTCTATA 879
DB 841 GAAAGCTGTGTTCCCGGGGCTCCCTTCTTTCTCTATA 880

RESULT 15
US-10-233-958-19
; Sequence 19. Application US/10233958
; Publication No. US20040009468A1
; GENERAL INFORMATION:
; APPLICANT: Mach, Bernard
; APPLICANT: Conrad, Bernard
; TITLE OF INVENTION: Allelic Variants of HER V-K18, Method for the Analysis
; TITLE OF INVENTION: Therapeutic and Use in the Determination of Genetic
; TITLE OF INVENTION: Predisposition for Disorders Involving the HERV-K18
; TITLE OF INVENTION: Proximus
; FILE REFERENCE: 23135-504
; CURRENT APPLICATION NUMBER: US/10/233,958
; PRIOR FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,513
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,522
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1010
; TYPE: DNA
; ORGANISM: Human endogenous retrovirus
US-10-233-958-19

Query Match 90.2%; Score 792.8; DB 15; Length 1010;
Best Local Similarity 94.5%; Pred. No. 8,6e-258;
Matches 832; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

QY 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTGTCTGTATAGAAAGATAGACAT 60
DB 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTGTCTGTATAGAAAGATAGACAT 60
QY 61 AGAGAGATCCATTTTGTATGTAAGTAAAGAAAAATCTTCTGCTTGAAGATTTCTTTATC 120
DB 61 AGAGAGATCCATTTTGTATGTAAGTAAAGAAAAATCTTCTGCTTGAAGATTTCTTTATC 120
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACAATGCTGTGTC -CACTCAGGGTTA 179
DB 121 TATGACCTTACCCCAACCCCGTCTCTGAAACAATGCTGTGTC -CACTCAGGGTTA 180
QY 180 AATGAGTTAAGGGGGTGGCAGAGATGCTGTTTAAACAATGCTTAAAGGCAAGCT 239
DB 181 AATGAGTTAAGGGGGTGGCAGAGATGCTGTTTAAACAATGCTTAAAGGCAAGCT 240
QY 240 CCTTAAGATCATCACCACCTCCCTAATCTCAAGTACCAGGACACAAAAATCGCGAAG 299
DB 241 CATTTAAGATCATCACCACCTCCCTAATCTCAAGTACCAGGACACAAAAATCGCGAAG 300
QY 300 GCGCAGAGACCTCTGCTAGGAAAGCAGATATGTCACACGTTTCTCCCATGTGATA 359
DB 301 GCTGAGAGGAGCTCTGCTAGGAAAGCAGATATGTCACACGTTTCTCCCATGTGATA 360
QY 360 GCGTGAATATGCTGCTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACACCCGT 419
DB 361 GCTGGAATATGCTGCTGGGAAAGGAAAGACCTGACCGTCCCGACCGGACACCCGT 420
QY 420 AAGAGGCTGCTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTSCAGTTGAGAC 479
DB 421 AAGAGGCTGCTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTSCAGTTGAGAC 480
QY 480 AAGAGAGGACATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
DB 481 AAGAGAGGACATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 540 CGATTGATGCTCCATCTTACTGAGATAGGAAAAACCGCTTGAAGGCTGGAGTGGAGC 599
DB 541 CGATTGATGCTCCATCTTACTGAGATAGGAAAAACCGCTTGAAGGCTGGAGTGGAGC 600
QY 600 TGGGGGAGAGAACTCTGCTTGTAAAGACACTGAGATGTTATGTAAGCATATCTAAA 659
DB 601 TGTGGGAGAGAACTCTGCTTGTAAAGACACTGAGATGTTATGTAAGCATATCTAAA 660
QY 660 GCACAGCACTTAATCTCTTACATCTGTCTATGATGCAAAAGACCTTGTTCACATGTTGTC 719
DB 661 GCACAGCACTTAATCTCTTACATCTGTCTATGATGCAAAAGACCTTGTTCACATGTTGTC 720
QY 720 TGTGACCCCTCTCCCAACATGCTTGTGACCTGACACATCCCTCTTCGAGAAACA 779
DB 721 TGTGACCCCTCTCCCAACATGCTTGTGACCTGACACATCCCTCTTCGAGAAACA 780
QY 780 CCACAGATATATCTAATAATCTAAGGAACTGAGAGCTGGGGAGTCCCTCATATGCT 839
DB 781 CCACAGATATATCTAATAATCTAAGGAACTGAGAGCTGGGGAGTCCCTCATATGCT 840
QY 840 GAAAGCTGTGTTCCCGGGGCTCCCTTCTTTCTCTATA 879
DB 841 GAAAGCTGTGTTCCCGGGGCTCCCTTCTTTCTCTATA 880

Search completed: February 28, 2004, 01:17:13

✓ Mon Mar 1 09:41:06 2004

Job time : 393.694 secs

us-10-016-604-5.rnpb

Page 11

This Page Blank (us019)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2004, 21:22:12 ; Search time 457.012 Seconds
(without alignments)
8170.833 million cell updates/sec

Title: US-10-016-604-5

Perfect score: 879

Sequence: 1 tctgggggaaagacagagag.....ccctctctctctctctcata 879

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_29Jan04:*
2: geneseqn1980s:*
3: geneseqn1980s:*
4: geneseqn2000s:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	879	100.0	879	6	ABX04572 Human end
2	879	100.0	879	6	ADCl16258 Human her
3	879	100.0	968	6	ABX04691 Human end
4	879	100.0	968	9	ADCl15768 LTR of he
5	856.6	97.5	32249	4	AA105336 Human rep
6	856.6	97.5	32249	4	ABL98205 Human tes
7	855	97.3	9179	6	ABX04612 Human end
8	851.8	96.9	968	6	ABX04693 Human end
9	851.8	96.9	968	6	ABX04612 Human end
10	845.4	96.2	35414	3	AAAD00147 TR12 tela
11	840.6	95.6	222930	6	ABK84349 Human CDN
12	837.4	95.3	968	6	ABX04695 Human end
13	837.4	95.3	968	6	ABX04694 Human end
14	837.4	95.3	968	6	ADCl15772 HML-2 LTR
15	837.4	95.3	968	6	ADCl15771 HML-2 LTR
16	835.4	95.0	6556	4	AAK70301 Human imm
17	834.4	94.9	96599	8	AA157703 Human pro
18	834.4	94.9	96599	8	ADA02672 Human PRL
19	834.4	94.9	96599	9	ADB72410 Human PRL
20	832.2	94.7	2431	4	AAK94873 Human ful
21	824.2	93.8	154902	6	ABQ88198 Human ost
22	813.6	92.6	139904	6	ABK83562 Human CDN
23	805.6	91.6	110000	6	ABX08336_11 Continuation (12 o

24	804	91.5	8095	4	AAK82357 Human imm
25	804	91.5	11122	6	ABX04611 Human end
26	797.6	90.7	975	7	ACC43225 Nucleotid
27	797.6	90.7	975	7	ACC43227 Nucleotid
28	797.6	90.7	2689	7	ACC43221 Nucleotid
29	797.6	90.7	2689	7	ACC43223 Nucleotid
30	797	90.7	1701	5	AA883303 DNA encod
31	797	90.7	1826	5	AA870436 DNA encod
32	796	90.6	972	7	ACC43230 Nucleotid
33	796	90.6	6020	7	ACC43236 Nucleotid
34	794.4	90.4	969	7	ACC43228 Nucleotid
35	794.4	90.4	975	7	ACC43226 Nucleotid
36	792.8	90.2	1010	7	ACC43229 Nucleotid
37	789.2	89.8	144792	9	ADCl15769 Human GPC
38	789.2	89.8	349981	9	ADCl15769 Human GPC
39	787.8	89.6	962	6	ABX04692 Human end
40	787.8	89.6	962	9	ADCl15769 Human end
41	784	89.2	198285	6	ABK84699 Human CDN
42	784	89.2	198285	6	ABN97319 Gene #381
43	778.6	88.6	1574	4	AA199316 Human exc
44	778.6	88.6	1574	5	AA163666 Human kid
45	766.2	87.2	1782	5	AA879381 DNA encod

ALIGNMENTS

RESULT 1
ID ABX04572 standard; DNA; 879 BP.
AC ABX04572;
DT 14-JAN-2003 (first entry)
XX Human endogenous retrovirus k (herv-k) U3R region.
XX Human endogenous retrovirus k (herv-k) U3R region.
XX Human; endogenous retrovirus; herv; prostate cancer; testicular cancer;
XX multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
XX cancer; transgenic animal; de.
XX Human endogenous retrovirus.
XX MO200246477-A2.
XX 13-JUN-2002.
XX 07-DEC-2001; 2001WO-US047824.
XX 07-DEC-2000; 2000US-0251830P.
XX 07-DEC-2001; 2001US-00016604.
XX (CHIR) CHIRON CORP.
XX Garcia P, Hardy SF, Williams LT, Escobedo J;
XX WPI; 2002-691475/74.
XX Novel isolated polypeptides useful for diagnosis of prostate cancer.
XX Claim 4; Page 129; 152pp; English.
XX The invention describes novel isolated polypeptides (I, Ib) useful for
XX diagnosing prostate cancer comprising obtaining a patient sample
XX containing prostate cells and detecting the presence or absence of an
XX expression product of a HML-2 endogenous retrovirus in a patient sample.
XX Polymorphic products associated with (I) are useful for diagnosis or treatment
XX of testicular cancer, multiple sclerosis or insulin-dependent diabetes
XX mellitus. An inhibitor of a HML-2 protease and a transdominant negative
XX mutant of HML-2 CORP are also useful in the manufacture of a medicament
XX for treating prostate cancer. (I) and (Ib) are useful for generating
XX antibodies specific to the polypeptides associated with cancer, as
XX targets for therapeutic intervention, and in immunising a transgenic

CC animal. This sequence represents a region of human endogenous retrovirus
CC (herv) of the HML-2 sub-group
XX
SQ Sequence 879 BP; 231 A; 212 C; 211 G; 225 T; 0 U; 0 Other;

Query Match 100.0%; Score 879; DB 6; Length 879;
Best Local Similarity 100.0%; Pred. No. 4.3e-288;
Matches 879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTCGTGTGTGAGAAAGAGATGATGATC 60
DB 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTCGTGTGTGAGAAAGAGATGATGATC 60
QY AGAGAGATCCATTTTGTATGTACTAAGAAAAATTTCTTCTCCCTTGAGATTTCTTATC 120
DB AGAGAGATCCATTTTGTATGTACTAAGAAAAATTTCTTCTCCCTTGAGATTTCTTATC 120
QY 121 TATGACCTTACCCCAACCCCGTCTCTGTGAAACATGTGCTGTGCTCACTGAGGTTAA 180
DB 121 TATGACCTTACCCCAACCCCGTCTCTGTGAAACATGTGCTGTGCTCACTGAGGTTAA 180
QY 181 ATGGATTAAAGGGCGGTGAGAGATGCTTGTGTTAAACAGATGCTTGAAGGAGATGCTC 240
DB 181 ATGGATTAAAGGGCGGTGAGAGATGCTTGTGTTAAACAGATGCTTGAAGGAGATGCTC 240
QY 241 CTTAAGAGTCATCAACCACTCCCTATCTCAAGTACCCAGGAGACAAAACTGCGGAAAG 300
DB 241 CTTAAGAGTCATCAACCACTCCCTATCTCAAGTACCCAGGAGACAAAACTGCGGAAAG 300
QY 301 CCGGAGGAGCTCTGCTCTAGGAAACCAAGTATGTTCCAACTTTTCCCATGTGATAG 360
DB 301 CCGGAGGAGCTCTGCTCTAGGAAACCAAGTATGTTCCAACTTTTCCCATGTGATAG 360
QY 361 CCGGAGGAGCTCTGCTCTAGGAAACCAAGTATGTTCCAACTTTTCCCATGTGATAG 420
DB 361 CCGGAGGAGCTCTGCTCTAGGAAACCAAGTATGTTCCAACTTTTCCCATGTGATAG 420
QY 421 AAGGTCGTGTCGTCAGAGAGATTTAGTAAAGAGAGAAATGCTCTTTCAGTTGAGACA 480
DB 421 AAGGTCGTGTCGTCAGAGAGATTTAGTAAAGAGAGAAATGCTCTTTCAGTTGAGACA 480
QY 481 AAGAGAAAGCATGCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 481 AAGAGAAAGCATGCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 GATTGTATGCTCCATCTAAGATAGTAAAGGAAACCCCTTAAGGCTGAGGTGGACT 600
DB 541 GATTGTATGCTCCATCTAAGATAGTAAAGGAAACCCCTTAAGGCTGAGGTGGACT 600
QY 601 GCGGCGAGCAATATGCTGTTGTGAAGCACTGAGATGTTATGTATGATCATATTAAG 660
DB 601 GCGGCGAGCAATATGCTGTTGTGAAGCACTGAGATGTTATGTATGATCATATTAAG 660
QY 661 CACAGCACTTAATCTTTTACATTTGTATGATGCAAAACCTTTGTTTCAATGTTTGTCT 720
DB 661 CACAGCACTTAATCTTTTACATTTGTATGATGCAAAACCTTTGTTTCAATGTTTGTCT 720
QY 721 GCTGACCTTCTCCCAACATTTGTGTGACCTTGAACATCCCTCTTTCGAGAAACAC 780
DB 721 GCTGACCTTCTCCCAACATTTGTGTGACCTTGAACATCCCTCTTTCGAGAAACAC 780
QY 781 CCAAGATGATCACTAATATCTAAGGAACTCAGAGGCTGAGGAGATCTTCATATGCTG 840
DB 781 CCAAGATGATCACTAATATCTAAGGAACTCAGAGGCTGAGGAGATCTTCATATGCTG 840
QY 841 AAGCTGTGCTCCCGGGTCCCTTCTTTCTTCTATA 879
DB 841 AAGCTGTGCTCCCGGGTCCCTTCTTTCTTCTATA 879
```

RESULT 2
ADCL6258
ID ADCL6258 standard; DNA; 879 BP.

```
XX AC ADCL6258;
XX 18-DEC-2003 (first entry)
XX DE Human herv-k(hml)-2(hom) U3R region nucleotide sequence SEQ ID NO:5.
XX KW prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
XX KM immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
XX KW multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.
XX OS Homo sapiens.
XX PN WO2003050258-A2.
XX PD 19-JUN-2003.
XX PE 09-DEC-2002; 2002MO-US039344.
XX PR 07-DEC-2001; 2001US-00016604.
XX PR 07-DEC-2001; 2001US-0340064P.
XX PR 07-DEC-2001; 2001MO-US047824.
XX PR 12-JUN-2002; 2002US-0388046P.
XX PA (CHIR ) CHIRON CORP.
XX PI Garcia P, Hardy SF, Williams LT, Escobedo J;
XX DR WPI; 2003-577296/54.
XX PT Diagnosing prostate cancer, comprises detecting the presence or absence
XX PT of HML-2 (human endogenous retrovirus) expression product in a patient
XX PT sample.
XX PS Disclosure; SEQ ID NO 5; 117bp; English.
XX PS
XX CC The present invention describes a method for diagnosing prostate cancer,
XX CC comprising detecting the presence or absence of HML-2 expression product
XX CC in a patient sample, where the expression product is produced by a
XX CC splicing event in which the 5' region and start codon of the env coding
XX CC region are joined to a downstream coding region in the reading frame +2
XX CC relative to that of env. Also described: (1) an isolated polynucleotide;
XX CC (2) an isolated polypeptide; (3) an antibody that binds to the
XX CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
XX CC sequences have cytosolic, neuroprotective, antidiabetic and
XX CC immunostimulant activities, and can be used in vaccines and in gene
XX CC therapy. A polynucleotide, polypeptide or antibody of the invention can
XX CC be used in the manufacture of a medicament for preventing or treating
XX CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
XX CC dependent diabetes mellitus. The present sequence represents a sequence
XX CC which is used in the exemplification of the present invention.
XX XX
SQ Sequence 879 BP; 231 A; 212 C; 211 G; 225 T; 0 U; 0 Other;
Query Match 100.0%; Score 879; DB 9; Length 879;
Best Local Similarity 100.0%; Pred. No. 4.3e-288;
Matches 879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTCGTGTGTGAGAAAGAGATGATGATC 60
DB 1 TGTGGGAAAAAGCAAGAGATGATGTTACTGTCGTGTGTGAGAAAGAGATGATGATC 60
QY 61 AGGAGATCCATTTTGTATGTACTAAGAAAAATTTCTTCTCCCTTGAGATTTCTTATC 120
DB 61 AGGAGATCCATTTTGTATGTACTAAGAAAAATTTCTTCTCCCTTGAGATTTCTTATC 120
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGTGCTCACTGAGGTTAA 180
DB 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGTGCTCACTGAGGTTAA 180
QY 181 ATGGATTAAAGGGCGGTGAGAGATGCTTGTGTTAAACAGATGCTTGAAGGAGATGCTC 240
DB 181 ATGGATTAAAGGGCGGTGAGAGATGCTTGTGTTAAACAGATGCTTGAAGGAGATGCTC 240
```

QY	241	CTTAAGAGTCATCACACATCCCTTAATCTCAAGTACCAGGAGACAAAACTGCGGAAG	3 00
Db	241	CTTAAGAGTCATCACATCCCTTAATCTCAAGTACCAGGAGACAAAACTGCGGAAG	3 00
QY	301	CCGACGGAGCCTCTGCTCTAAGAAAAGCCAGGTATTGTCCAACGTTTCTCCCATGTATAG	3 60
Db	301	CCGACGGAGCCTCTGCTCTAAGAAAAGCCAGGTATTGTCCAACGTTTCTCCCATGTATAG	3 60
QY	361	CCTGAAAATTGGCCCTCTGTGGGAAAGGAAAGACTGACCGTCCCCCAGCCCGACACCCGTA	4 20
Db	361	CCTGAAAATTGGCCCTCTGTGGGAAAGGAAAGACTGACCGTCCCCCAGCCCGACACCCGTA	4 20
QY	421	AAGGCTCTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGCAAGTTGAGACA	4 80
Db	421	AAGGCTCTGTGCTGAGAGAGATTAGTAAAGAGAGAAATGCTCTTGCAAGTTGAGACA	4 80
QY	481	AGAGAGAGGCATCTGTCTCTGCTCTGCTCCCTGGGACATGAAATGTCTCGGTATPAAACC	5 40
Db	481	AGAGAGAGGCATCTGTCTCTGCTCTGCTCCCTGGGACATGAAATGTCTCGGTATPAAACC	5 40
QY	541	GATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGGCTGAGAGTGGACCT	6 00
Db	541	GATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGGCTGAGAGTGGACCT	6 00
QY	601	GCGGCGACCAATACTGCTCTTGTAAAGCACTGAGATTTATGTATGCTATCTAAAG	6 60
Db	601	GCGGCGACCAATACTGCTCTTGTAAAGCACTGAGATTTATGTATGCTATCTAAAG	6 60
QY	661	CACAGCATCTTAATCCTTTACATTTGCTATGATGCAAAAGCCTTTGTTACATGTTTGTCT	7 20
Db	661	CACAGCATCTTAATCCTTTACATTTGCTATGATGCAAAAGCCTTTGTTACATGTTTGTCT	7 20
QY	721	GCTGACCTCTCCCAACAATTGCTCTGTGACCTGACACATCCCGCTTCCTGAGAAACAC	7 80
Db	721	GCTGACCTCTCCCAACAATTGCTCTGTGACCTGACACATCCCGCTTCCTGAGAAACAC	7 80
QY	781	CCACAGATGATCATGTAATACTAAGGAACTGAGAGGCTGCGGGATCTCCATATGCTG	8 40
Db	781	CCACAGATGATCATGTAATACTAAGGAACTGAGAGGCTGCGGGATCTCCATATGCTG	8 40
QY	841	AAAGCTGGTTCCCGGGATCCCTTTCTTTCTTTCTCTATA	8 79
Db	841	AAAGCTGGTTCCCGGGATCCCTTTCTTTCTTTCTCTATA	8 79

	RESULT 3
ABX04691	ID ABX04691 standard; DNA; 968 BP.
XX	
AC	ABX04691;
XX	
DT	14-JAN-2003 (first entry)
DE	Human endogenous retrovirus k (herv-k) long terminal repeat #1.
XX	
KW	Human; endogenous retrovirus; herv; prostate cancer; testicular cancer; multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease; cancer; transgenic animal; de.
KM	
OS	Human endogenous retrovirus.
XX	
PN	WO200246477-A2.
XX	
PD	13-JUN-2002.
PF	07-DEC-2001; 2001WO-US047824.
XX	
PR	07-DEC-2000; 2000US-0251830P.
XX	
PA	07-DEC-2001; 2001US-00016604.
XX	
CHIR	(CHIR) CHIRON CORP.

PI Garcia P, Hardy SF, Williams LT, Escobedo J,
DR WPI; 2002-691475/74.
XX
XX Novel isolated polypeptides useful for diagnosis of prostate cancer.
PT
PS Claim 18; Page 148; 152pp; English.
XX
XX The invention describes novel isolated polypeptides (I, Ib) useful for
CC diagnosing prostate cancer comprising obtaining a patient sample
CC containing prostate cells and detecting the presence or absence of an
CC expression product of a HmL-2 endogenous retrovirus in a patient sample.
CC Polynucleotides associated with (I) are useful for diagnosis or treatment
CC of testicular cancer, multiple sclerosis or insulin-dependent diabetes
CC mellitus. An inhibitor of a HmL-2 protease and a transdominant negative
CC mutant of HmL-2 CORF are also useful in the manufacture of a medicament
CC for treating prostate cancer. (I) and (Ib) are useful for generating
CC antibodies specific to the polypeptides associated with cancer, as
CC targets for therapeutic intervention, and in immunising a transgenic
CC animal. This sequence represents a region of human endogenous retrovirus
CC (heLV) of the HmL-2 sub-group
CC
CC
SQ Sequence 968 BP; 249 A; 242 C; 225 G; 252 T; 0 U; 0 Other;

Query Match 100.0%; Score 879; DB 6; Length 968;
Best Local Similarity 100.0%; Pred. No. 4.Ee-288;
Matches 879; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGTGGGGAAGCAAGAAGATCATGTTACTGTGTCTGTGTGAAGAAGTAGACAT 60
DB 1 TGTGGGGAAGCAAGAAGATCATGTTACTGTGTGTGTGTGAAGAAGTAGACAT 60
OY 61 AGAGACTCCATTGTTTATGTATGTAAGAAAATTCCTCGCCTTGAGATTCTGTAAATC 120
DB 61 AGGAGACTCCATTGTTTATGTATGTAAGAAAATTCCTCGCCTTGAGATTCTGTAAATC 120
OY 121 TATGACCTTACCCTCCCAACCCCGTGCTCTGTAAACATGTGCTGTGTCCAATCAGGGTTAA 180
DB 121 TATGACCTTACCCTCCCAACCCCGTGCTCTGTAAACATGTGCTGTGTCCAATCAGGGTTAA 180
OY 121 TATGACCTTACCCTCCCAACCCCGTGCTCTGTAAACATGTGCTGTGTCCAATCAGGGTTAA 180
DB 121 TATGACCTTACCCTCCCAACCCCGTGCTCTGTAAACATGTGCTGTGTCCAATCAGGGTTAA 180
OY 181 ATGATTTAAGGCGGCGGTGAGAGATGTCGTTTTTAAACAGATGCTTGAAGGAGCATGCTC 240
DB 181 ATGATTTAAGGCGGCGGTGAGAGATGTCGTTTTTAAACAGATGCTTGAAGGAGCATGCTC 240
OY 241 CTTAAGAGTCAATCACACTCCCTAATCTCAAGTACCCAGGACAATAAACTGGGGAAG 300
DB 241 CTTAAGAGTCAATCACACTCCCTAATCTCAAGTACCCAGGACAATAAACTGGGGAAG 300
OY 241 CTTAAGAGTCAATCACACTCCCTAATCTCAAGTACCCAGGACAATAAACTGGGGAAG 300
DB 241 CTTAAGAGTCAATCACACTCCCTAATCTCAAGTACCCAGGACAATAAACTGGGGAAG 300
OY 301 CGCAGAGGACCTTGCTGCCTTAGGAAAACGAGTATTTGCCAGGTTTCTCCCATGTGATAG 360
DB 301 CGCAGAGGACCTTGCTGCCTTAGGAAAACGAGTATTTGCCAGGTTTCTCCCATGTGATAG 360
OY 301 CGCAGAGGACCTTGCTGCCTTAGGAAAACGAGTATTTGCCAGGTTTCTCCCATGTGATAG 360
DB 301 CGCAGAGGACCTTGCTGCCTTAGGAAAACGAGTATTTGCCAGGTTTCTCCCATGTGATAG 360
OY 361 CCTGAATATATGCGCTTCGTGGGAAGGAAAGACTGACCGTCCCCAGCCGACACCCGTA 420
DB 361 CCTGAATATATGCGCTTCGTGGGAAGGAAAGACTGACCGTCCCCAGCCGACACCCGTA 420
OY 421 AAGGATCTGTCTAGAGAGATTGTAAAAAGAAAGGAATGCGCTCTTGACAGTTGAGACA 480
DB 421 AAGGATCTGTCTAGAGAGATTGTAAAAAGAAAGGAATGCGCTCTTGACAGTTGAGACA 480
OY 481 AGAGAAAGGACCTGTCTCTCCTGCTCCCTGAGCAATGGAATGCTTGGTATTAACCC 540
DB 481 AGAGAAAGGACCTGTCTCTCCTGCTCCCTGAGCAATGGAATGCTTGGTATTAACCC 540
OY 541 GATTGTATGCTCCATCTAATGAGTAAAGGAAACCCGCTTAAGGCTGGAAGTGGACCT 600
DB 541 GATTGTATGCTCCATCTAATGAGTAAAGGAAACCCGCTTAAGGCTGGAAGTGGACCT 600
OY 601 GGCGGAGCAATATGCTTTGTAAAGCACTGATGTTTATGTGTATGCAATATCTAAAG 660
DB 601 GGCGGAGCAATATGCTTTGTAAAGCACTGATGTTTATGTGTATGCAATATCTAAAG 660
OY 661 CACGACCTTATCTTATCTATGATGCAAGAACCTTTGTTCACATGTTTGTCT 720

DE Human reproductive system related antigen DNA SEQ ID NO: 8024.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001339.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209457P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 05-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.

PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236377P.
PR 29-SEP-2000; 2000US-0236378P.
PR 29-SEP-2000; 2000US-0236387P.
PR 29-SEP-2000; 2000US-0236389P.
PR 29-SEP-2000; 2000US-0236399P.
PR 29-SEP-2000; 2000US-0236399P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239357P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251039P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251858P.
PR 08-DEC-2000; 2000US-0251859P.
PR 08-DEC-2000; 2000US-0251869P.

PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX PT used in preventing, treating or ameliorating a medical condition.
XX
XX
XX Disclosure; SEQ ID NO 8024; 1297bp + Sequence Listing; English.
XX
XX
XX The present invention provides the protein and coding sequences of a
XX CC number of human reproductive system related antigens. These can be used
XX CC in the prevention and treatment of reproductive system disorders,
XX CC including cancer. The present sequence is a genomic sequence encoding a
XX CC protein of the invention
XX
XX
XX Sequence 32249 BP; 10716 A; 6390 C; 5851 G; 9292 T; 0 U; 0 Other;

Query Match 97.5%; Score 856.6; DB 4; Length 32249;
Best Local Similarity 98.4%; Pred. No. 1.6e-279;
Matches 865; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTGTAAGAAAGATGACAT 60
DB 25971 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTGTAAGAAAGATGACAT 26030
QY 61 AGGAGACTCATTGTTGTAATGTAAGAAATAATCTTCTGCTTGAAGATTCTGTTATC 120
DB 26031 AGGAGACTCATTGTTGTAATGTAAGAAATAATCTTCTGCTTGAAGATTCTGTTATC 26090
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCTTCACTCAGGGTTAA 180
DB 26091 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCTTCACTCAGGGTTAA 26150
QY 181 ATGGATTAAAGGCGGTGAGAGATGCTTGTAAACAGATGCTTGAAGGACAGATCTC 240
DB 26151 ATGGATTAAAGGCGGTGAGAGATGCTTGTAAACAGATGCTTGAAGGACAGATCTC 26210
QY 241 CTTAAGATCATCAACACTCCCTAATCTCAAGTACCCAGGAGACAAATACTGGGAAG 300
DB 26211 CTTAAGATCATCAACACTCCCTAATCTCAAGTACCCAGGAGACAAATACTGGGAAG 26270
QY 301 CCGGAGGAGCTCTGCTTAGAAAGCAAGTATTGTCCACAGTTTCTCCCATGTGATAG 360
DB 26271 CCGGAGGAGCTCTGCTTAGAAAGCAAGTATTGTCCACAGTTTCTCCCATGTGATAG 26330
QY 361 CTTGAATATGAGCTCTGTGGGAAGGAAAGCCTGACGTCCTCCGACCCGAGACCCGTA 420
DB 26331 CTTGAATATGAGCTCTGTGGGAAGGAAAGCCTGACGTCCTCCGACCCGAGACCCGTA 26390
QY 421 AAGGCTCTGCTGAGAGAGATTAGTAAAGAGAGAGAAATGCTCTTGGAGTTGAGACA 480
DB 26391 AAGGCTCTGCTGAGAGAGATTAGTAAAGAGAGAGAAATGCTCTTGGAGTTGAGACA 26450
QY 481 AGAGGAAGGATCTGTCTCTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 26451 AGAGGAAGGATCTGTCTCTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 26510
QY 541 GATTGTATGCTCACTACTGAGATGAGTAAAGCACTGAGATGTTATGTTATGATCTAAAG 600
DB 26511 GATTGTATGCTCACTACTGAGATGAGTAAAGCACTGAGATGTTATGTTATGATCTAAAG 26570
QY 601 GCGGCGACATACTGCTTTGTAAAGCACTGAGATGTTATGTTATGATCTAAAG 660
DB 26571 GCGGCGACATACTGCTTTGTAAAGCACTGAGATGTTATGTTATGATCTAAAG 26630
QY 661 CACGACCTTAATCTTTACATTTGTATGATGCAAAAGCCTTGTTCACATGTTTGTCT 720
DB 26631 CACGACCTTAATCTTTACATTTGTATGATGCAAAAGCCTTGTTCACATGTTTGTCT 26690

DB 26631 CACGACCTTAATCTTTACATTTGTATGATGCAAAAGCCTTGTTCACATGTTTGTCT 26690
QY 721 GCTGACCTCTCCCAACAATTGTTGTGACCCCTGACACATCCCTCTTGAAGAACAC 780
DB 26691 GCTGACCTCTCCCAACAATTGTTGTGACCCCTGACACATCCCTCTTGAAGAACAC 26750
QY 781 CCACAGATGATCTGTAATATCTAAGGAACTCAGAGAGCTGCGGGAATCTTCCATATGCTG 840
DB 26751 CCACAGATGATCTGTAATATCTAAGGAACTCAGAGAGCTGCGGGAATCTTCCATATGCTG 26810
QY 841 AACGCTGTTCCCGGGATCCCTTCTTCTCTATA 879
DB 26811 AACGCTGTTCCCGGGATCCCTTCTTCTCTATA 26849
RESULT 6
ID ABL98205 standard; DNA; 32249 BP.
XX
XX ABL98205;
XX
XX
XX 21-JUN-2002 (first entry)
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2857.
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX KM reproductive system disorder; urinary system disorder; gene therapy;
XX KM cardiovascular disorder; respiratory disorder; neurological disorder;
XX KM gastrointestinal disease; infection; cytostatic; gene; ds.
XX
XX Homo sapiens.
XX
XX WO200155317-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001329.
XX
XX
XX 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.
XX PR 14-JUL-2000; 2000US-0218290P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 26-JUL-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225213P.
XX PR 14-AUG-2000; 2000US-0225214P.
XX PR 14-AUG-2000; 2000US-0225266P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225757P.
XX PR 14-AUG-2000; 2000US-0225758P.
XX PR 14-AUG-2000; 2000US-0225759P.
XX PR 18-AUG-2000; 2000US-0226279P.
XX PR 22-AUG-2000; 2000US-0226681P.
XX PR 22-AUG-2000; 2000US-0226688P.
XX PR 22-AUG-2000; 2000US-0227182P.
XX PR 23-AUG-2000; 2000US-0227009P.

PD 19-JUN-2003.
 XX 09-DEC-2002; 2002MO-US039344.
 XX 07-DEC-2001; 2001US-00016604.
 PR 07-DEC-2001; 2001US-034064P.
 PR 07-DEC-2001; 2001MO-US047824.
 PR 12-JUN-2002; 2002US-0388046P.
 XX
 PA (CHIR) CHIRON CORP.
 PI Garcia P, Hardy SF, Williams LT, Escobedo J;
 XX WPI; 2003-577296/54.
 DR
 XX
 PT Diagnosing prostate cancer, comprises detecting the presence or absence
 PT of HML-2 (human endogenous retrovirus) expression product in a patient
 PT sample.
 PS Disclosure; SEQ ID NO 46; 117bp; English.
 XX
 XX The present invention describes a method for diagnosing prostate cancer,
 CC comprising detecting the presence or absence of HML-2 expression product
 CC in a patient sample, where the expression product is produced by a
 CC splicing event in which the 5' region and start codon of the env coding
 CC region are joined to a downstream coding region in the reading frame +2
 CC relative to that of env. Also described: (1) an isolated polynucleotide;
 CC (2) an isolated polypeptide; (3) an antibody that binds to the
 CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
 CC sequences have cytostatic, neuroprotective, antidiabetic and
 CC immunostimulant activities, and can be used in vaccines and in gene
 CC therapy. A polynucleotide, polypeptide or antibody of the invention can
 CC be used in the manufacture of a medicament for preventing or treating
 CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
 CC dependent diabetes mellitus. The present sequence represents a sequence
 CC which is used in the exemplification of the present invention.
 XX
 XX
 SQ Sequence 968 BP; 248 A; 238 C; 226 G; 256 T; 0 U; 0 Other;
 Query Match 96.9%; Score 851.8; DB 9; Length 968;
 Best Local Similarity 98.1%; Pred. No. 8.4e-279;
 Matches 862; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

DB 421 AAGGCTCTGTCTGAGAGGATTTAGTATAGAGAAAGCATGCTCTTGCAGTTGAGACA 480
 QY 481 AGAGGAAGCATCTGTCTCTGCTGTCTCTGGGCAATGGAATGTCTGGTATAAACC 540
 DB 481 AGAGGAAGCATCTGTCTCTGCTGTCTCTGGGCAATGGAATGTCTGGTATAAACC 540
 QY 541 GATTGTATGTCTCATCTAATGAGATAGGAAAAACCGCTTAGGGCTGAGGGGACT 600
 DB 541 GATTGTATGTCTCATCTAATGAGATAGGAAAAACCGCTTAGGGCTGAGGGGACT 600
 QY 601 GCGGAGAGCAATCTCTTTGTAAGCACTGAGATGTTATGTATGATATCTAAAG 660
 DB 601 GCGGAGAGCAATCTCTCTTTGTAAGCACTGAGATGTTATGTATGATATCTAAAG 660
 QY 661 CACAGCACTTAATCTTTACATTTGTATGATGCAAAAGACCTTTGTACATGTTGCT 720
 DB 661 CACAGCACTTAATCTTTACATTTGTATGATGCAAAAGACCTTTGTACATGTTGCT 720
 QY 721 GCTGACCTCTCCCAATTTGTTGTGACCTGACATCTCCCTCTTGAGAAACAC 780
 DB 721 GCTGACCTCTCCCAATTTGTTGTGACCTGACATCTCCCTCTTGAGAAACAC 780
 QY 781 CCAAGATGATGATTAATCTAAGGAACTGAGAGGCTGGGGGATCCCTCATATGCTG 840
 DB 781 CCAAGATGATGATTAATCTAAGGAACTGAGAGGCTGGGGGATCCCTCATATGCTG 840
 QY 841 AAGCGTGTCTCCCGGATCTCTTCTTTCTCTATA 879
 DB 841 AAGCGTGTCTCCCGGATCTCTTCTTTCTCTATA 879
 RESULT 10
 AAD00147
 ID AAD00147 standard; DNA; 35414 BP.
 AC AAD00147;
 XX
 XX 31-JUL-2000 (first entry)
 XX
 DE TR12 related DNA-11.
 XX
 KW Human tumour necrosis factor receptor related gene 12; TR12; cytostatic;
 KW anti-infective; anti-human immunodeficiency virus; nephrotoxic; cancer;
 KW antiarteriosclerotic; analgesic; cardiatic; immunosuppressant; virucide;
 KW anti-allergic; anti-inflammatory; hepatocytic; antidiabetic; vulnery;
 KW anti-asthmatic; antiproliferative; antibacterial; antitumor; angiogenic;
 KW microbial infection; nephritis; bone disease; pain; atherosclerosis;
 KW cardiovascular disorder; allergy; inflammation; asthma; cardiomyopathy;
 KW neurodegenerative disease; Alzheimer's disease; graft rejection;
 KW liver disease; autoimmune disease; diabetes; influenza; psoriasis;
 KW glomerulonephritis; septic shock; ulcerative colitis; angiogenesis;
 KW hematompleisis; wound healing; ds.
 KW
 OS Homo sapiens.
 XX
 PN WO200023572-A1.
 XX
 PD 27-APR-2000.
 XX
 PF 19-OCT-1999; 99MO-US024413.
 XX
 PR 20-OCT-1998; 98US-0104950P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ni J, Ruben SM;
 XX
 DR WPI; 2000-339675/29.
 XX
 PT Tumor necrosis factor receptor related gene 12 (TR12) polynucleotides and
 PT polypeptides, useful for prevention, treatment, and diagnosis of various
 PT diseases such as cancer, atherosclerosis, pain, and cardiovascular
 PT disorders.

XX Disclosure; Page 227-246; 250bp; English.

PS The patent discloses the use of tumour necrosis factor receptor related

XX gene 12 (TR12). TR12 gene and protein are useful for treatment of

CC diseases including, cancer, microbial infections (including human

CC immunodeficiency virus (HIV)), nephritis, bone diseases, atherosclerosis,

CC pain, cardiovascular disorders, acquired immune deficiency syndrome

CC (AIDS), allergy, inflammation, neurodegenerative diseases (e.g.

CC Alzheimer's, Parkinson's), graft rejection, liver disease, autoimmune

CC disease, cardiomyopathy, diabetes, influenza, asthma, psoriasis,

CC glomerulonephritis, septic shock and ulcerative colitis. They may also be

CC useful for promoting angiogenesis, regulating haematopoiesis, wound

CC healing, and has various diagnostic and research applications. The

CC present sequence is a specifically related polynucleotide sequence that

CC is specifically excluded from the scope of the present invention

XX

50 Sequence 35414 BP; 8874 A; 8376 C; 8358 G; 9806 T; 0 U; 0 Other;

Query Match 96.2%; Score 845.4; DB 3; Length 35414;

Best Local Similarity 97.6%; Pred. No. 1.1e-275;

Matches 858; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TGTGGGAAAAGCAAGAGATCATGATGTTTCTGTCTGTGTAAGAAAGTAGACAT 60

DB 8681 TGTGGGAAAAGCAAGAGATCATGATGTTTCTGTCTGTGTAAGAAAGTAGACAT 8740

QY 61 AGGAGACTCCATTTTGTATGATTAAGAAAATTTCTGCGCTTGAGATTCTGTATC 120

DB 8741 AGGAGACTCCATTTTGTATGATTAAGAAAATTTCTGCGCTTGAGATTCTGTATC 8800

QY 121 TATGACCTATCCCAACCCCGTCTCTGTAACATGCTGTGCTCACTCAGAGTTAA 180

DB 8801 TATGACCTATCCCAACCCCGTCTCTGTAACATGCTGTGCTCACTCAGAGTTAA 8860

QY 181 ATGATTAAGGGCGGTGAGAGATGCTTTGTTAAACAGATGCTTGAAGGAGCATGCTC 240

DB 8861 ATGATTAAGGGCGGTGAGAGATGCTTTGTTAAACAGATGCTTGAAGGAGCATGCTC 8920

QY 241 CTTAAGAGTATCAACCACTCCCTTAATCTCAAGTACCCAGGACACAAAATCTGGGAAAG 300

DB 8921 CTTAAGAGTATCAACCACTCCCTTAATCTCAAGTACCCAGGACACAAAATCTGGGAAAG 8980

QY 301 CCGGAGGAGCCTGCTGCTAGGAAAGCAGATGTTGTCACGTTCTCCCATGATGATG 360

DB 8981 CCGGAGGAGCCTGCTGCTAGGAAAGCAGATGTTGTCACGTTCTCCCATGATGATG 9040

QY 361 CTTGAAATATGCTCTGCTGAGGAAAGAGCTGACCGTCCCGACCCGACACCCGTA 420

DB 9041 TCTGAAATATGCTCTGCTGAGGAAAGAGCTGACCGTCCCGACCCGACACCCGTA 9100

QY 421 AAGGGTCTGCTGAGAGATTAAGTAAAGAGAGAGATGCTCTTTCAGTTGAGCA 480

DB 9101 AAGGGTCTGCTGAGAGATTAAGTAAAGAGAGAGATGCTCTTTCAGTTGAGCA 9160

QY 481 AGAGGAGGAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

DB 9161 AGAGGAGGAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9220

QY 541 GATTGTATGCTCATCTACTAGATAGGAAAGAAACCCGCTTGAAGGCTGAGGAGCT 600

DB 9221 GATTGTATGCTCATCTACTAGATAGGAAAGAAACCCGCTTGAAGGCTGAGGAGCT 9280

QY 601 GCGGGGAGCAATGCTGCTTTGTAAGCACTGAGATGTTATGTTATGATATCTAAAG 660

DB 9281 GCGGGGAGCAATGCTGCTTTGTAAGCACTGAGATGTTATGTTATGATATCTAAAG 9340

QY 661 CACAGCACTTAATCTTATGATGTTATGTTATGTTATGTTATGTTATGTTATGTTATG 720

DB 9341 CACAGCACTTAATCTTATGATGTTATGTTATGTTATGTTATGTTATGTTATGTTATG 9400

QY 721 GCTGACCTCTCCCAATGTTGTTGAGACCTGAGACATCCCTCTTTCGAGAAACAC 780

DB 9401 GCTGACCTCTCCCAATGTTGTTGAGACCTGAGACATCCCTCTTTCGAGAAACAC 9460

QY 781 CCACAGATGATCAGTAAATATCTAAGGAACTCAGAGGCTGGGAGATCTCATATGCTG 840

DB 9461 CCACAGATGATCAGTAAATATCTAAGGAACTCAGAGGCTGGGAGATCTCATATGCTA 9520

QY 841 AACGCTGTTCCCGGGTCCCTCTTCTTCTCTATA 879

DB 9521 AACGCTGTTCCCGGGGCCCCCTTATTTCTTCTATA 9559

RESULT 11

ABK84349/C

ID ABK84349 standard; cDNA: 222930 BP.

XX

AC ABK84349;

XX

DT 14-AUG-2002 (first entry)

XX

DE Human cDNA differentially expressed in granulocytic cells #920.

XX

KM Human; se; granulocytic cell; DNA chip; bacterial infection;

KM viral infection; parasitic infection; protozoal infection;

KM fungal infection; sterile inflammatory disease; psoriasis;

KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;

KM cardiac reperfusion injury; renal reperfusion injury; ARDS;

KM adult respiratory distress syndrome; inflammatory bowel disease;

KM Crohn's disease; ulcerative colitis; periodontal disease;

KM granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.

XX

XX WO200228999-A2.

XX

XX 11-APR-2002.

XX

XX 03-OCT-2001; 2001WO-US030821.

XX

XX 03-OCT-2000; 2000US-0237189P.

PR

PA (GENE-) GENE LOGIC INC.

XX

PI Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;

XX

DR WP1; 2002-435328/46.

XX

PT Detecting granulocyte activation by detecting differential expression of

PT genes associated with granulocyte activation, which serves as diagnostic

PT markers that is useful for monitoring disease states and drug toxicity.

XX

PS Claim 1; SEQ ID NO 920; 114bp; English.

XX

XX The invention relates to detecting (M1) granulocyte (GC) activation

CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by

CC DNA chip analysis as given in the specification, and comparing the

CC expression level to an expression level in an unactivated GC, where

CC differential expression of Gs is indicative of GCA. Also included are

CC modulating (M2) GA by contacting GC with an agent that alters the

CC expression of at least one gene in Gs; (2) screening (M3) for an agent

CC capable of modulating GCA or an inflammation (especially chronic) in a

CC tissue, an allergic response in a subject, exposure of a subject to a

CC pathogen or sterile inflammatory disease using the gene expression

CC profile; (3) detecting (M4) an inflammation (especially chronic) in a

CC tissue, an allergic response in a subject, exposure of a subject to a

CC pathogen or sterile inflammatory disease, by detecting the level of

CC expression in a sample of the tissue of gene(s) from Gs, where the level

CC of expression of the gene is indicative of inflammation; (4) treating

CC (M5) an inflammation (especially chronic) or in a tissue, an allergic

CC response in a subject, exposure of a subject to a pathogen or sterile

CC inflammatory disease, by contacting a tissue having inflammation with an

CC agent that modulates the expression of gene(s) from Gs in the tissue. M1

CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful

CC for screening an agent capable of modulating GCA preferably in an

Db 121 TATTAACCTTACCCCAACCCCGTCTCTGTAAACATGTCTATGTCAACTCAAGATTGA 180
Qy 181 ATGATTAAGGCGGTCAGAGATGTCTTTTAAACAGATGCTTGAAGGAGCATGTCTC 240
Db 181 ATGATTAAGGCGGTCAGAGATGTCTTTTAAACAGATGCTTGAAGGAGCATGTCTC 240
Qy 241 CTTAAGAGTCATCAACCTCCCTAATCTCAAGTACCCAGGACACAAAACCTGGGAAAG 300
Db 241 CTTAAGAGTCATCAACCTCCCTAATCTCAAGTACCCAGGACACAAAACCTGGGAAAG 300
Qy 301 CCGGAGGACCTCTGCTAGGAAAGCCAGATTTGTCCAAAGTTTCTCCCATGTGATAG 360
Db 301 CCGGAGGACCTCTGCTAGGAAAGCCAGATTTGTCCAAAGTTTCTCCCATGTGATAG 360
Qy 361 CTTGAATATATGCTCTGTGGAAAGGAAAGCTGACCGTCCCGACCCGACACCGGTA 420
Db 361 TCTGAATATATGCTCTGTGGAAAGGAAAGCTGACCGTCCCGACCCGACACCGGTA 420
Qy 421 AAGGCTCTGTGCTAGAGAGATTAGTAAAGAGAGAAATGCTTGGAGTTGAGACA 480
Db 421 AAGGCTCTGTGCTAGAGAGATTAGTAAAGAGAGAAATGCTTGGAGTTGAGACA 480
Qy 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qy 541 GATTGTATGCTCCATCTACTAGATAGGAGAAACCCGCTTGAAGGCTGGAGCTGGA 600
Db 541 GATTGTATGCTCCATCTACTAGATAGGAGAAACCCGCTTGAAGGCTGGAGCTGGA 600
Qy 601 GCGGGGAGCAATATGCTTTGTAAAGCACTGAGATGTTATGTATGATGATCTTAAAG 660
Db 601 GCGGGGAGCAATATGCTTTGTAAAGCACTGAGATGTTATGTATGATGATCTTAAAG 660
Qy 661 CACAGCACTTAATCTTATCTATGATGATGATGATGATGATGATGATGATGATGATG 720
Db 661 CACAGCACTTAATCTTATCTATGATGATGATGATGATGATGATGATGATGATGATG 720
Qy 721 GCTGACCTCTCTCCCAACATTTGTTGACCTGACATCCCTCTTGAAGAAC 780
Db 721 GCTGACCTCTCTCCCAACATTTGTTGACCTGACATCCCTCTTGAAGAAC 780
Qy 781 CCAAGATGATCACTAATAATACTAAGGAACCTGAGGCTGGGAGTCTTCAATGCTG 840
Db 781 CCAAGATGATCACTAATAATACTAAGGAACCTGAGGCTGGGAGTCTTCAATGCTG 840
Qy 841 AAGGCTGTTCCCGGGTCCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 879
Db 841 AAGGCTGTTCCCGGGTCCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 879

RESULT 13
ABX04694
ID ABX04694 standard; DNA; 968 BP.
AC ABX04694;
XX
XX 14-JUN-2003 (first entry)
XX DE Human endogenous retrovirus k (herv-k) long terminal repeat #4.
XX DE Human endogenous retrovirus; herv; prostate cancer; testicular cancer;
XX KW multiple sclerosis; insulin-dependent diabetes mellitus; HML-2 protease;
XX KW cancer; transgenic animal; ds.
XX OS Human endogenous retrovirus.
XX PN W020024647-A2.
XX PD 13-JUN-2002.
XX PF 07-DEC-2001; 2001WO-US047824.

XX 07-DEC-2001; 2000US-0251830P.
PR 07-DEC-2001; 2001US-00016604.
XX (CHIR) CHIRON CORP.
XX Garcia P, Hardy SF, Williams LT, Escobedo J;
XX WPI; 2002-691475/74.
XX Novel isolated polypeptides useful for diagnosis of prostate cancer.
XX Claim 18; Page 149; 1522p; English.
XX The invention describes novel isolated polypeptides (I, Ib) useful for
XX diagnosing prostate cancer comprising obtaining a patient sample
XX containing prostate cells and detecting the presence or absence of an
XX expression product of a HML-2 endogenous retrovirus in a patient sample.
XX Polynucleotides associated with (I) are useful for diagnosis or treatment
XX of testicular cancer, multiple sclerosis or insulin-dependent diabetes
XX mellitus. An inhibitor of a HML-2 protease and a transdominant negative
XX mutant of HML-2 CORP are also useful in the manufacture of a medicament
XX for treating prostate cancer. (I) and (Ib) are useful for generating
XX antibodies specific to the polypeptides associated with cancer, as
XX targets for therapeutic intervention, and in immunising a transgenic
XX animal. This sequence represents a region of human endogenous retrovirus
XX (herv) of the HML-2 sub-group
XX
XX Sequence 968 BP; 252 A; 238 C; 224 G; 254 T; 0 U; 0 Other;
XX
XX Query Match 95.3%; Score 837.4; DB 6; Length 968;
XX Best Local Similarity 97.0%; Pred. No. 6.8e-274;
XX Matches 853; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
Qy 1 TGTGGGAGAAAGCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Db 1 TGTGGGAGAAAGCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Qy 61 AGAGAGCTCCATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 120
Db 61 AGAGAGCTCCATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 120
Qy 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGCTGATCAGGAGTTAA 180
Db 121 TATGACCTTACCCCAACCCCGTCTCTTAAACATGTGCTGATCACTCAAGATTAA 180
Qy 181 ATGATTAAGGCGGTCAGAGATGTCTTTTAAACAGATGCTTGAAGGAGCATGTCTC 240
Db 181 ATGATTAAGGCGGTCAGAGATGTCTTTTAAACAGATGCTTGAAGGAGCATGTCTC 240
Qy 241 CTTAAGAGTCATCAACCTCCCTAATCTCAAGTACCCAGGACACAAAACCTGGGAAAG 300
Db 241 CTTAAGAGTCATCAACCTCCCTAATCTCAAGTACCCAGGACACAAAACCTGGGAAAG 300
Qy 301 CCGGAGGACCTCTGCTAGGAAAGCCAGATTTGTCCAAAGTTTCTCCCATGTGATAG 360
Db 301 CCGGAGGACCTCTGCTAGGAAAGCCAGATTTGTCCAAAGTTTCTCCCATGTGATAG 360
Qy 361 CTTGAATATATGCTCTGTGGAAAGGAAAGCTGACCGTCCCGACCCGACACCGGTA 420
Db 361 TCTGAATATATGCTCTGTGGAAAGGAAAGCTGACCGTCCCGACCCGACACCGGTA 420
Qy 421 AAGGCTCTGTGCTAGAGAGATTAGTAAAGAGAGAAATGCTTGGAGTTGAGACA 480
Db 421 AAGGCTCTGTGCTAGAGAGATTAGTAAAGAGAGAAATGCTTGGAGTTGAGACA 480
Qy 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 481 AGAGGAAGGATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qy 541 GATTGTATGCTCCATCTACTAGATAGGAGAAACCCGCTTGAAGGCTGGAGCTGGA 600
Db 541 GATTGTATGCTCCATCTACTAGATAGGAGAAACCCGCTTGAAGGCTGGAGCTGGA 600

QY 601 GCGGAGCAATAGTCTTTGTAAAGACATGATGTTATGTATGATATCTAAAG 660
CC
CC 601 GCGGAGCAATAGTCTTTGTAAAGACATGATGTTATGTATGATATCTAAAG 660
XX
QY 661 CACAGCACTTAATCTTACCTTGTATGATGAAAGACCTTGTTCACATGTTTCT 720
CC
CC 661 CACAGCACTTAATCTTGTATGATGAAAGACCTTGTTCACATGTTTCT 720
XX
Db 661 CACAGCACTTAATCTTGTATGATGAAAGACCTTGTTCACATGTTTCT 720
QY 721 GCTGACCTCTCCCAATGTTGTGACCTGACACATCCCTCTTCGAGAAACAC 780
CC
CC 721 GCTGACCTCTCCCAATGTTGTGACCTGACACATCCCTCTTCGAGAAACAC 780
Db 721 GCTGACCTCTCCCAATGTTGTGACCTGACACATCCCTCTTCGAGAAACAC 780
QY 781 CCACAGATGATCAATAATACTAAGGAATCTGAGGCTGGCGGATCTTCATATGCTG 840
CC
CC 781 CCACAGATGATCAATAATACTAAGGAATCTGAGGCTGGCGGATCTTCATATGCTG 840
Db 781 CCACAGATGATCAATAATACTAAGGAATCTGAGGCTGGCGGATCTTCATATGCTG 840
QY 841 AAGCGTGTCCCGGGTCCCTCTTCTTCTCTATA 879
CC
CC 841 AAGCGTGTCCCGGGTCCCTCTTCTTCTCTATA 879
Db 841 AAGCGTGTCCCGGGTCCCTCTTCTTCTCTATA 879

RESULT 14
ADCL5772
ID ADCL5772 standard, DNA, 968 BP.

XX
AC ADCL5772;

DT 18-DEC-2003 (first entry)

DE HML-2 LTR nucleotide sequence SEQ ID NO:48.

XX prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
KM immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
KM multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.

XX Homo sapiens.

XX WO2003050258-A2.

XX 19-JUN-2003.

PF 09-DEC-2002; 2002WO-US039344.

XX 07-DEC-2001; 2001US-00016604.

PR 07-DEC-2001; 2001US-034064P.

PR 07-DEC-2001; 2001WO-US047824.

PR 12-JUN-2002; 2002US-0388046P.

XX (CHIR) CHIRON CORP.

PI Garcia P, Hardy SF, Williams LT, Escobedo J;

DR WPI; 2003-577296/54.

XX Diagnosing prostate cancer, comprises detecting the presence or absence
PT of HML-2 (human endogenous retrovirus) expression product in a patient
XX sample.

PS Disclosure; SEQ ID NO 48; 117bp; English.

XX The present invention describes a method for diagnosing prostate cancer,
CC comprising detecting the presence or absence of HML-2 expression product
CC in a patient sample, where the expression product is produced by a
CC splicing event in which the 5' region and start codon of the env coding
CC region are joined to a downstream coding region in the reading frame +2
CC relative to that of env. Also described: (1) an isolated polynucleotide;
CC (2) an isolated polypeptide; (3) an antibody that binds to the
CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
CC sequences have cytosolic, neuroprotective, antidiabetic and
CC immunostimulant activities, and can be used in vaccines and in gene
CC therapy. A polynucleotide, polypeptide or antibody of the invention can
CC be used in the manufacture of a medicament for preventing or treating

CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
CC dependent diabetes mellitus. The present sequence represents a sequence
CC which is used in the exemplification of the present invention.
XX

SO Sequence 968 BP; 250 A; 237 C; 225 G; 256 T; 0 U; 0 Other;

Query Match 95.3%; Score 837.4; DB 9; Length 968;
Best Local Similarity 97.0%; Pred. No. 6.8e-274;
Matches 853; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 TGTGGGAAAAGCAAGAGATGATGATGTTTACTGTGTCTGTGTGAAAAGATGACAT 60
Db 1 TGTGGGAAAAGCAAGAGATGATGATGTTTACTGTGTCTGTGTGAAAAGATGACAT 60
QY 61 AGGAGATCCATTTTGTATGATGATGAAAATTTCTTGTGCTTGAATTTCTTATC 120
Db 61 AGGAGATCCATTTTGTATGATGATGAAAATTTCTTGTGCTTGAATTTCTTATC 120
QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAAACATGCTGTGTCACTCAGGTTAA 180
Db 121 TATGACCTTACCCCAACCCCGTCTCTGAAAACATGCTGTGTCACTCAGGTTAA 180
QY 181 ATGATTTAAGGGGAGTGCAGATGTGCTTGTAAACAGATGCTTGAAGGACATCTC 240
Db 181 ATGATTTAAGGGGAGTGCAGATGTGCTTGTAAACAGATGCTTGAAGGACATCTC 240
QY 241 CTTAAGATCATCAACATCTCCTAATCTCAAGTACCCAGGACACAAAATCTGGAAG 300
Db 241 CTTAAGATCATCAACATCTCCTAATCTCAAGTACCCAGGACACAAAATCTGGAAG 300
QY 301 CCGCAGGACCTCTGCTTGAAGAAAGCAGATTTTGTCAACGTTTCTCCCATGTATAG 360
Db 301 CCGCAGGACCTCTGCTTGAAGAAAGCAGATTTTGTCAACGTTTCTCCCATGTATAG 360
QY 361 CTTGAATATGAGCTCGTGGGAAAGGAAAGCCTGACCTGCAAGGTTTCTCCCATGTATAG 420
Db 361 CTTGAATATGAGCTCGTGGGAAAGGAAAGCCTGACCTGCAAGGTTTCTCCCATGTATAG 420
QY 421 AAGGCTGTGCTGAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTGAAGTTGAGACA 480
Db 421 AAGGCTGTGCTGAGAGAGATTAGTAAAGAGAAAGAAATGCTCTTGAAGTTGAGACA 480
QY 481 AGAGAAAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 481 AGAGAAAGCATCTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 GATTGATGCTCCATCTACTAGATGAGGAAAGAAACGCTTAGGGCTGAGGTGGACT 600
Db 541 GATTGATGCTCCATCTACTAGATGAGGAAAGAAACGCTTAGGGCTGAGGTGGACT 600
QY 601 GCGGAGCAATATCTGTTTGTAAAGCACTGAGATGTTTATGTGTATGATATCTAAAG 660
Db 601 GCGGAGCAATATCTGTTTGTAAAGCACTGAGATGTTTATGTGTATGATATCTAAAG 660
QY 661 CACAGCACTTAATCTTATCATTTGTATGATGAAAGACCTTGTTCACATGTTTCT 720
Db 661 CACAGCACTTAATCTTATCATTTGTATGATGAAAGACCTTGTTCACATGTTTCT 720
QY 721 GCTGACCTCTCCCAATTTGTGTCCTTGTGACCTGACACATCCCTCTTCGAGAAACAC 780
Db 721 GCTGACCTCTCCCAATTTGTGTCCTTGTGACCTGACACATCCCTCTTCGAGAAACAC 780
QY 781 CCACAGATGATCAATAATACTAAGGAATCTGAGGCTGGCGGATCTTCATATGCTG 840
Db 781 CCACAGATGATCAATAATACTAAGGAATCTGAGGCTGGCGGATCTTCATATGCTG 840
QY 841 AAGCGTGTCCCGGGTCCCTTCTTCTTCTATA 879
Db 841 AAGCGTGTCCCGGGTCCCTTCTTCTTCTATA 879

RESULT 15
ADCL5771

ID ADCL5771 standard; DNA; 968 BP.
 AC ADCL5771;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE HML-2 LTR nucleotide sequence SEQ ID NO:47.
 XX
 KM prostate cancer; HML-2; env; cytosolic; neuroprotective; antidiabetic;
 KM immunostimulant; vaccine; gene therapy; breast cancer; testicular cancer;
 KM multiple sclerosis; insulin-dependent diabetes mellitus; human; gene; ds.
 OS Homo sapiens.
 XX
 PN WO2003050258-A2.
 XX
 PD 19-JUN-2003.
 XX
 PF 09-DEC-2002; 2002WO-US039344.
 XX
 PR 07-DEC-2001; 2001US-00016604.
 PR 07-DEC-2001; 2001US-0340664P.
 PR 07-DEC-2001; 2001WO-US047824.
 PR 12-JUN-2002; 2002US-0388046P.
 XX
 PA (CHIR) CHIRON CORP.
 PI Garcia P, Hardy SF, Williams LT, Escobedo J;
 XX
 DR WPI; 2003-577236/54.
 XX
 PT Diagnosing prostate cancer, comprises detecting the presence or absence
 PT of HML-2 (human endogenous retrovirus) expression product in a patient
 PT sample.
 XX
 PS Disclosure; SEQ ID NO 47; 117bp; English.
 XX
 CC The present invention describes a method for diagnosing prostate cancer,
 CC comprising detecting the presence or absence of HML-2 expression product
 CC in a patient sample, where the expression product is produced by a
 CC splicing event in which the 5' region and start codon of the env coding
 CC region are joined to a downstream coding region in the reading frame +2
 CC relative to that of env. Also described: (1) an isolated polynucleotide;
 CC (2) an isolated polypeptide; (3) an antibody that binds to the
 CC polypeptide; and (4) an immunogenic composition. Prostate cancer related
 CC sequences have cytosolic, neuroprotective, antidiabetic and
 CC immunostimulant activities, and can be used in vaccines and in gene
 CC therapy. A polynucleotide, polypeptide or antibody of the invention can
 CC be used in the manufacture of a medicament for preventing or treating
 CC prostate, breast or testicular cancer, multiple sclerosis or insulin-
 CC dependent diabetes mellitus. The present sequence represents a sequence
 CC which is used in the exemplification of the present invention.
 XX
 SQ Sequence 968 BP; 252 A; 238 C; 224 G; 254 T; 0 U; 0 Other;

Query Match 95.3%; Score 837.4; DB 9; Length 968;
 Best Local Similarity 97.0%; Pred. No. 6.8e-274;
 Matches 853; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTGAGAAAGATGACAT 60
 DB 1 TGTGGGAAAAGCAAGAGATCAGATTGTTACTGTCTGTGTGAGAAAGATGACAT 60
 QY 61 AGGAGACTTCATTTTGTATGACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
 DB 61 AGGAGACTTCATTTTGTATGACTAAGAAAATTTCTTGCTTGAGATTCTGTATC 120
 QY 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCCACTCAGGGTTAA 180
 DB 121 TATGACCTTACCCCAACCCCGTCTCTGAAACATGTGTGTCCACTCAGGGTTAA 180
 QY 181 ATGATTAAGGGCGGTGAGATGTGCTTTGTTAAAGATGCTGAAGCAGCATGCTC 240
 DB 181 ATGATTAAGGGCGGTGAGATGTGCTTTGTTAAAGATGCTGAAGCAGCATGCTC 240

DB 181 ATGATTAAGGGCGGTGAGATGTGCTTTGTTAAAGATGCTGAAGCAGCATGCTC 240
 QY 241 CTTAAGATCATACCACTCCCTTAATCTCAAGTACCCAGGAGACAAAACCTGGGAGG 300
 DB 241 CTTAAGATCATACCACTCCCTTAATCTCAAGTACCCAGGAGACAAAACCTGGGAGG 300
 QY 301 CCGAGGAGACTCTGCTCTAGAAAAGCCAGTATTTGCAACGTTTCTCCCATGTGATAG 360
 DB 301 CCGAGGAGACTCTGCTCTAGAAAAGCCAGTATTTGCAACGTTTCTCCCATGTGATAG 360
 QY 361 CTTAGAGGACTCTGCTCTAGAAAAGCCAGTATTTGCAACGTTTCTCCCATGTGATAG 420
 DB 361 CTTAGAGGACTCTGCTCTAGAAAAGCCAGTATTTGCAACGTTTCTCCCATGTGATAG 420
 QY 421 AAGGCTGTGCTGAGAGAGATTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 DB 421 AAGGCTGTGCTGAGAGAGATTAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 QY 481 AGAGAGAGAGATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 DB 481 AGAGAGAGAGATCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 QY 541 GATTGATGCTCATCTACTAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 DB 541 GATTGATGCTCATCTACTAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 QY 601 GCGGAGAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
 DB 601 GCGGAGAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
 QY 661 CACAGCACTTAATCTCTTATCATTTGATGATGCAAGACCTTGTTCATATGTTTGTCT 720
 DB 661 CACAGCACTTAATCTCTTATCATTTGATGATGCAAGACCTTGTTCATATGTTTGTCT 720
 QY 721 GCTGACCTCTCCCAATTTGTTGTAACCTGACACATCTCCCTCTTGAGAGAAACAC 780
 DB 721 GCTGACCTCTCCCAATTTGTTGTAACCTGACACATCTCCCTCTTGAGAGAAACAC 780
 QY 781 CCACAGATGATCACTAATTAATCTAAGGAACTCAGAGGCTGCGGAGATCTCCATATGCTG 840
 DB 781 CCACAGATGATCACTAATTAATCTAAGGAACTCAGAGGCTGCGGAGATCTCCATATGCTG 840
 QY 841 AAGCTGTTTCCCGGGTCCCTTCTTCTTCTCTATA 879
 DB 841 AAGCTGTTTCCCGGGTCCCTTCTTCTTCTCTATA 879

Search completed: February 27, 2004, 22:44:02
 Job time : 465.012 secs

THIS PAGE BLANK (03/07)